(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04596 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/21354

(22) International Filing Date: 5 July 2001 (05.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/216,723	7 July 2000 (07.07.2000)	US
09/812,633	19 March 2001 (19.03.2001)	US
09/812,471	19 March 2001 (19.03.2001)	US

(71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). **THE BRIGHAM AND WOMEN'S HOSPITAL, INC.** [US/US]; 75 Francis Street, Boston, MA 02115 (US).

(72) Inventors: **BENJAMIN, Thomas, L.**; 595 Putnam Avenue, Cambridge, MA 02139 (US). **LI, Dawei**; 16 Copenger Street, Apartment #1, Boston, MA 02120 (US). **MOK, Samuel, C.**; 125 Pleasant Street, #607, Brookline, MA 02446 (US). **CRAMER, Daniel, W.**; 151 Shaw Road, Chestnut Hill, MA 02167 (US). **MA, Yupo**; 24 West Street, Sharon, MA 02067 (US).

(74) Agent: **BIEKER-BRADY, Kristina**; Clark & Elbing LLP, 176 Federal Street, Boston, MA 0211-2214 (US).

(81) Designated States *(national)*: CA, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

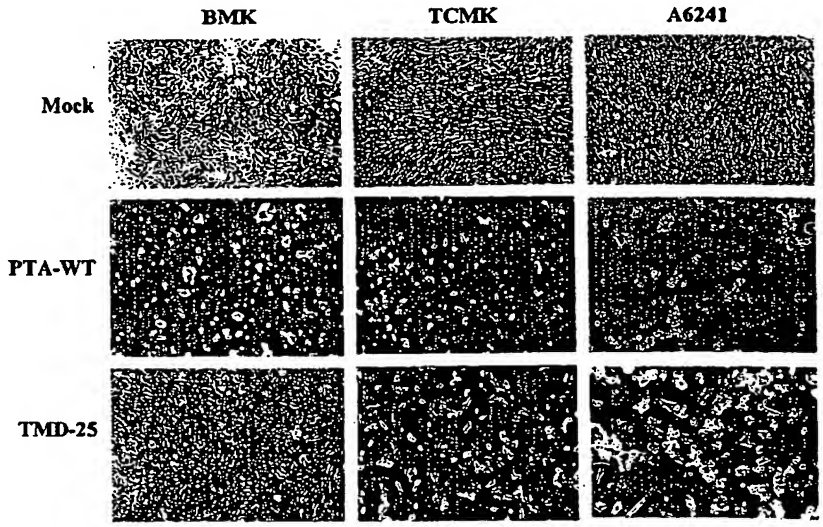
Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DIAGNOSING AND TREATING CANCER CELLS USING MUTANT VIRUSES



(57) Abstract: The invention provides methods for the identification of genes and their encoded proteins involved in the susceptibility to proliferative disorders, including cancer, using a tumor host range mutant virus (T-HR mutant). In addition, the invention provides methods for the diagnosis of abnormally proliferating cells in a subject, using a T-HR mutant. The invention also features, T-HR mutants that can be used to kill cancer cells such as ones carrying a Sal2 alteration. Furthermore, the invention features the analysis of Sal2 nucleic acids and proteins for diagnosing and treating patients having proliferative disorders, including cancer, involving mutations in a *Sal2* gene and encoded protein. Also encompassed by the invention are transgenic and knockout mice including Sal2 nucleic acids or proteins and mutants thereof.

WO 02/04596 A2

DIAGNOSING AND TREATING CANCER CELLS USING MUTANT VIRUSES

Statement as to Federally Sponsored Research

The present research was supported by a grant from the National Cancer Institute (Number R35 CA44343). The U.S. government has certain rights to this invention.

Field of the Invention

The field of the invention is regulation of cellular proliferation.

Background of the Invention

Transforming genes of DNA tumor viruses perform essential functions in virus growth, acting largely as proto-oncogene activators or tumor suppressor gene inactivators. The isolation and characterization of mutant viruses that are able to propagate in cells containing a mutation in known proto-oncogene or tumor suppressor genes has been useful in identifying and studying the viral equivalents or interactors of these genes. The transforming gene of the highly oncogenic murine polyoma virus was identified through studies of host range mutants isolated using polyoma transformed 3T3 cells as the permissive host and normal 3T3 cells as the non-permissive host. This approach requires expression a known viral protein by the permissive host, since it is based on the idea of complementation between cell-associated wild-type viral genes and an infecting virus mutant. In addition to its use with polyoma virus, the complementation approach has also been successfully used with other oncogenic DNA viruses, e.g., with 293 cells expressing adenovirus E1A/E1B genes and COS cells expressing the SV40 large T antigen. Complementing cell lines have also been used in other systems to propagate specifically defective virus mutants for vaccine development and other purposes. However, by design, these types of

systems rely on permissive hosts constructed with known gain-of-function mutations and are only applicable to mutants in known viral genes, as well as to viruses with known mutations, since the host cell must express a functional version of the mutant viral protein.

The use of mutant adenoviruses unable to inactivate p53 or the retinoblastoma protein (pRb) to kill cancer cells lacking one of these proteins has been previously described (Patent Nos. U.S. 5,677,178 and WO 94/18992). It was well known prior to these observations that these two genes are mutated in a variety of cancers.

While a number of genes are known to be involved in the progression towards cancer, there is a significant need for the development of a general, unbiased method for identifying new genes involved in the pre-disposition for, or progression of cancer or other proliferative disorders. Furthermore, methods for diagnosing and treating patients with mutations in known as well as newly identified genes would greatly aid in the management of cancer.

Summary of the Invention

The invention features novel tumor host range viruses, for identifying mammalian cancer susceptibility genes, such as tumor suppressor genes and proto-oncogenes, and methods for diagnosing and treating patients having proliferative disorders, such as cancers, involving mutations in such genes. Furthermore, the invention features the use of *Sal2* nucleic acids and proteins in methods of identifying a mammal having, or at risk of acquiring, a proliferative disease.

The tumor host range mutant viruses (T-HR mutants) used in the methods of the invention contain mutations that prevent the virus from propagating in normal cells. These viruses are, however, able to propagate in abnormally proliferating cells because of genetic changes that are present in these cells, such as the inactivation of tumor suppressor genes or the activation of proto-oncogenes. A T-HR mutant that infects a normal cell is unable to propagate in such a cell because it is unable to inactivate a tumor suppressor gene or to

activate a proto-oncogene due to a mutation in the viral genome. In contrast, if this T-HR mutant infects an abnormally proliferating cell that already has a tumor suppressor gene inactivated, this virus is able to propagate. Likewise, if such a T-HR mutant infects an abnormally proliferating cell that contains an activated proto-oncogene, the virus is also able to propagate.

The Tumor Host Range Mutant System

Since a T-HR mutant is unable to propagate in normal cells, but is able to propagate in abnormally proliferating cells, the first aspect of the invention features a method of using T-HR mutants to identify a cellular protein that is involved in the susceptibility to cancer and other proliferative disorders. This method involves: (a) infecting a normal cell and an abnormally proliferating cell with a collection of uncharacterized mutant viruses; (b) identifying a mutant virus from the collection that can grow in an abnormally proliferating cell and can not grow in a normal cell (i.e., a T-HR mutant); (c) identifying the mutated viral gene or mutated protein in the virus, where this mutation allows the virus to grow on the abnormally proliferating cell; and (d) screening to identify the cellular proteins which interact with the wild-type viral protein, but not with the mutated protein.

In a preferred embodiment of the above aspect of the invention, the abnormally proliferating cell infected with the collection of uncharacterized mutant viruses is also uncharacterized. In an additional preferred embodiment, the cellular and viral proteins can be identified by, for example, using an assay that detects protein-protein interactions (e.g., a GST-pull-down assay). These proteins may be, for example, tumor suppressor proteins or proto-oncogene products; however the retinoblastoma tumor suppressor protein and the gene encoding this protein are specifically excluded from this and all other aspects of the invention. In another preferred embodiment, the method of this aspect is used to isolate a mutant virus (i.e., a T-HR mutant).

Preferred viruses with a mammalian, preferably human, host range used in this and other aspects of the invention include, for example, simian virus 40,

human polyoma virus, parnovirus, papilloma virus, herpes virus, and primate adenoviruses.

The second aspect of the invention features a method of determining the presence or absence of an alteration in the genetic material of a cell that involves determining whether such a cell can act as a permissive host for the growth of a characterized T-HR mutant, where the T-HR mutant is capable of propagating in an abnormally proliferating cell and not capable of propagating in a normal cell. The retinoblastoma and p53 genes are specifically excluded from this aspect of the invention.

In a preferred embodiment of the above aspect of the invention, the alteration of the genetic material to be tested for in the cell indicates that the organism carrying this alteration is at an increased risk of developing a proliferative disease. Preferably, this genetic alteration is in a tumor suppressor gene or in a proto-oncogene. In another preferred embodiment, the T-HR mutant has been characterized as being complemented by a mutation in a specific tumor suppressor or proto-oncogene. In an additional preferred embodiment of the above aspects of the invention, the cells used in the methods of the invention are from a mammal, for example, a human.

In another aspect, the invention features a method of killing a cell with a proliferative disease that involves: (i) contacting a cell with a proliferative disease, for example, a mammalian cell, with a T-HR mutant; and (ii) allowing the T-HR mutant to lyse this cell. In a preferred embodiment, the TH-R mutant may be a T-HR mutant specific for a cell carrying a *Sal2* mutation, for example, the TMD-25 T-HR mutant virus. In an additional preferred embodiment of this aspect, the mammalian cell is from a human. The mammalian cell may also be in a mammal, for example a human, with a proliferative disorder. In a further embodiment, the T-HR mutant may be administered, for example, in a pharmaceutically acceptable carrier. In addition, the T-HR mutant may be administered, for example, by parenteral, intravenous, intraperitoneal, intramuscular, subcutaneous, or subdermal injection. The T-HR mutant, however, may also be administered orally, nasally, topically, or as an aerosol.

The Use of Sal2 as a Diagnostic and Treatment Tool

A further aspect of the invention features a method of identifying a mammal having, or at increased risk of acquiring, a proliferative disease. This method includes determining whether there is a proliferative disease-associated alteration in a *Sal2* nucleic acid of the mammal. An example of a proliferative disease-associated Sal2 alteration is the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1. In one embodiment, the method is used to identify a mammal, preferably a human, having a proliferative disease, while in another embodiment, the method is used to identify a mammal at increased risk of acquiring a proliferative disease.

In another embodiment of this aspect of the invention, determining whether the mammal has or is at increased risk of acquiring a proliferative disease is done by, for example, polymerase chain reaction (PCR) amplification, single nucleotide polymorphism (SNP) determination, restriction fragment length polymorphism (RFLP) analysis, hybridization analysis, or mismatch detection analysis.

In addition, identifying the alteration may also involve: (i) contacting a first nucleic acid probe which is specific for binding to the human *Sal2* nucleic acid containing the alteration with a nucleic acid from a cell from the mammal under conditions which allow the first nucleic acid probe to anneal to complementary sequences in the cell; and (ii) detecting duplex formation between the first nucleic acid probe and the complementary sequences. The nucleic acid probe of step (i), which is, for example, at least 12 contiguous nucleotides in length, may be derived from the human *Sal2* nucleic acid containing a proliferative disease-associated alteration. The cell may be from a physiological sample, which may contain, for example, mRNA or the nucleic acid probe of step (i) annealed to the mRNA. Furthermore, another embodiment of this aspect includes a second nucleic acid probe, where the first and second nucleic acid probes are PCR primers, and where the human *Sal2* nucleic acid or a fragment is amplified using PCR between steps (i) and (ii).

In another embodiment of this aspect, the cell may be selected from a physiological sample, for example, containing abnormally proliferating or normal tissue, and may be from human tissue, blood, ovarian tissue, bladder tissue, colon tissue, and cells grown in culture.

An additional aspect of the invention features a method of identifying a mammal having, or at increased risk of acquiring, a proliferative disease involving, determining whether there is an alteration in a Sal2 protein of the mammal. Preferably the method is used to identify a mammal, such as a human, having a proliferative disease, or at increased risk of acquiring a proliferative disease. In preferred embodiments of this method an antibody specific for either the human, or for a proliferative disease-associated mutant Sal2 protein is used.

A further aspect of the invention encompasses a knockout mouse featuring a knockout mutation in a genomic *mSal2* gene. This knockout mouse may also contain, for example, a nucleic acid construct including a mutant *Sal2* gene and this mutant *Sal2* gene may be conditionally expressed. In a preferred embodiment, the mutant *Sal2* gene, for example a human *Sal2* gene, encodes a protein that contains a substitution of a Cys for the Ser at position 73 of SEQ ID NO:1. However, the Sal2 protein may also be wild-type.

An additional aspect of the invention features a transgenic mouse whose genome includes a nucleic acid construct that contains a *Sal2* nucleic acid, which is operably linked to transcriptional regulatory elements and encodes a Sal2 protein, e.g., a mutant Sal2 protein. The mutant Sal2 protein may also be a human Sal2 protein, for example, one that has a modification of function. Furthermore, the transgenic mouse may contain a mouse Sal2 protein, e.g., the protein of SEQ ID NO:3. This mouse Sal2 protein may be mutant, such as a mouse Sal2 protein containing the substitution of a Cys for the Ser at position 73 of SEQ ID NO:3.

In preferred embodiments of this aspect, the transcriptional regulatory elements include a promoter that is a tissue-specific promoter, such that the nucleic acid is expressed, and the protein is produced at detectable levels, in cells selected from the group consisting of ovarian, bladder, and colon cells.

However, the transcriptional regulatory element may also include the wild-type *Sal2* promoter.

In a further embodiment of this aspect, the transgenic mouse develops ovarian tumors, and these tumors may metastasize. The invention also includes a cell line, such as an ovarian cell line, derived from cells isolated from the transgenic mouse.

An additional aspect, the invention encompasses a method of identifying a compound which alters cell proliferation, the method involving: a) contacting a first cell with a test compound, and b) measuring whether the test compound alters proliferation in the first cell, relative to a second cell not contacted with the test compound, wherein the first and second cells have a proliferative disease-associated alteration in a *Sal2* nucleic acid. In a preferred embodiment of this aspect, the ability of the test compound to alter proliferation is determined by measuring the ability of a virus, for example, a T-HR mutant virus, to propagate in the first cell contacted with the test compound, relative to the second cell not contacted with the test compound. In addition, the first and second cells may be mammalian cells, for example, human cells. Furthermore, these cells may be ovarian cells. The cells may also be in the same mammal or in different mammals and the mammal may be a transgenic mouse or a knockout mouse containing a knockout mutation in a genomic *mSal2* gene.

A final aspect of the invention features a method of identifying a compound which alters cell proliferation, the method involving: a) exposing a cell or a cell extract to a test compound, and b) measuring whether the test compound alters Sal2 levels, for example, Sal2 protein or nucleic acid levels, relative to Sal2 levels in a cell or cell extract not exposed to the test compound. In a preferred embodiment of this aspect, the cell has a proliferative disease-associated alteration in a *Sal2* nucleic acid or the extract is from a cell having a proliferative disease-associated alteration in a *Sal2* nucleic acid. This cell or cell extract may be mammalian, e.g., human. Furthermore, the cell may be in a mammal, for example, a transgenic mouse whose genome includes a nucleic acid construct containing a *Sal2* nucleic acid, which is operably linked to

transcriptional regulatory elements and encodes a Sal2 protein, or a knockout mouse comprising a knockout mutation in a genomic *mSal2* gene. In another embodiment, the exposing in step a) of this aspect is done with a cell and this cell is an ovarian cell.

Definitions

"Tumor host range mutant virus (T-HR mutant)," as used herein, refers to a virus that has a reduced ability to replicate and disseminate in a normal cell, relative to the replication of a wild-type virus in the same type of cell, but is able to replicate and disseminate in a cell having abnormal proliferation. The abnormally proliferating cell may, for example, have one or more mutations in a gene or genes involved in the regulation of cell growth, of the cell cycle, or of programmed cell death (e.g., apoptosis). These genes include, for example, tumor suppressor genes and proto-oncogenes, but any cellular gene that a virus must inactive or activate in order to grow is also included. Adenoviruses having mutations in the p53 and retinoblastoma genes are specifically excluded.

Reference herein to a "collection of uncharacterized mutant viruses" refers to a sample of viruses where at least one of the viruses, in a collection of at least 1000 viruses, (e.g., 0.1%) carries at least one mutation in at least one of the genes of the viral genome. Preferably, at least 10%, 25%, 30%, or 50% of the viruses in this collection carry at least one mutation in at least one of the genes in the viral genome. In addition, such mutations preferably inactivate viral proteins that are necessary for transforming a host cell into a cancer cell. The types of mutations that may be present in the viral genes include, for example, point mutations, deletions, insertions, duplications, and inversions. Furthermore, the mutations may result in modification of function, such as a partial or a complete loss-of-function of the viral gene. Preferably the virus has a mammalian host range (e.g., rodent or primate), most preferably a human host range. Viruses that may be used in such a collection include, for example, simian virus 40, human polyoma virus, parnovirus, papilloma virus, herpes virus, and primate adenoviruses. However, any virus that needs to overcome a cell cycle checkpoint

or affect a signal transduction pathway in order to propagate may be used in this collection.

"T-HR mutant specific for a Sal2 mutation," as used herein, refers to a TH-R mutant virus that is able to propagate in a cell containing a genetic alteration in a *Sal2* gene. For example, the "T-HR mutant specific for a Sal2 mutation" may be the TMD-25 T-HR mutant virus described herein.

"Sal2 mutation", as used herein, refers to a genetic change in the nucleic acid sequence of a Sal2 gene, for example, SEQ ID NO:2 and SEQ ID NO:4, which results in the abnormal proliferation, or predisposition to abnormal proliferation, of a cell carrying such a change. Preferably, the genetic change is a missense mutation. Most preferably, the mutation is a substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

"Uncharacterized abnormally proliferating cell," as used herein, refers to a cell where the cause of the abnormal proliferation is unknown. For example, the genetic alteration that results in abnormal proliferation has not been identified. However, other features of the cell may be characterized.

"Cancer susceptibility gene," as used herein, refers to any gene that, when altered, increases the likelihood that the organism carrying the gene will develop a proliferative disorder during its lifetime. Examples of such genes include proto-oncogenes, tumor suppressor genes, and genes involved in the regulation of cell growth, the cell cycle, and apoptosis.

"Proliferative disease," as used herein, refers to any genetic change within a differentiated cell that results in the abnormal proliferation of a cell. Such changes include mutations in genes involved in the regulation of the cell cycle, of growth control, or of apoptosis and can further include tumor suppressor genes and proto-oncogenes. Specific examples of proliferative diseases are the various types of cancer, such as ovarian cancer. However, proliferative diseases may also be the result of the cell becoming infected with a transforming virus.

"Proliferative disease-associated alteration," as used herein, refers to any genetic change within a differentiated cell that results in the abnormal proliferation of a cell. Preferably, such a genetic change correlates with a

statistically significant (e.g., the p-value is less than or equal to 0.05) increase in the risk of acquiring a proliferative disease. Examples of such genetic changes include mutations in genes involved in the regulation of the cell cycle, of growth control, or of apoptosis and can further include mutations in tumor suppressor genes and proto-oncogenes. A further example of a proliferative disease-associated alteration is the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

"Abnormal proliferation," as used herein, refers to a cell undergoing cell division that normally does not undergo cell division or that undergoes cell division at an increased frequency when compared to a wild-type cell.

The term "alteration," when used herein, in reference to a gene, refers to a change in the nucleic acid sequence. Such a change may include, for example, insertions, deletions, and substitutions of one or more nucleic acids, as well as inversions and duplications.

"Genetic lesion," as used herein, refers to a nucleic acid change. Examples of such a change include single nucleic acid changes as well as deletions and insertions of one or more nucleic acid. However, genetic lesions can also include duplications and inversions. In addition, a genetic lesion may be a naturally-occurring polymorphism, for example, one that predisposes an organism carrying the polymorphism to acquiring a proliferative disease.

"Polymorphism," as used herein, refers to an alteration in a nucleic acid sequence, for example, a gene. Such an alteration may result in a codon change, which in turn may result in, for example, the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

"Modification of function," as used herein, refers to a change in the function of the protein. Such a change can, for example, result in the partial or complete loss of function, but it can also result in a gain of function.

As used herein, the term "promoter" is intended to encompass transcriptional regulatory elements, that is, all of the elements that promote or regulate transcription, including core elements required for basic interactions

between RNA polymerase, transcription factors, upstream elements, enhancers, and response elements.

"Operably linked," as referred to herein, describes the functional relationship between nucleic acid sequences, for example, a promoter sequence, and a gene to be expressed. Operably linked nucleic acids may be part of a contiguous sequence. However a physical link is not necessary for two nucleic acid sequences to be operably linked. For example, enhancers can exert their effect over long distances and therefore do not require a physical link in sequence to the gene whose transcription they affect.

Reference herein to the "transcriptional regulatory elements" of a gene or a class of genes includes both the entire gene as well as an intact region of naturally-occurring transcriptional regulatory elements. Also included are transcription regulatory elements modified by, for example, rearrangement of the elements, deletion of some elements or of extraneous sequences, and insertion of heterologous elements.

The term "knockout," as used herein, refers to an alteration in the sequence of a specific gene that results in a decrease of function of that gene. Preferably the alteration results in undetectable or insignificant expression of the gene and in a complete or partial loss of function. Furthermore, the disruption may be conditional, e.g., dependent on the presence of tetracycline. Knockout animals may be homozygous or heterozygous for the gene of interest. In addition, the term knockout includes conditional knockouts, where the alteration of the target gene can occur, for example, as a result of exposure of the animal to a substance that promotes target gene alteration, introduction of an enzyme that promotes recombination at the target gene site (e.g., Cre in the Cre-lox system, or FLP in the FLP/FRT system), or any other method for directing target gene alteration.

"Conditionally expressed," as used herein, refers to any method that may be used to control expression of a gene, such as a transgene. These methods may, for example, include the use of promoters that are regulated by a substance, such as tetracycline, that can be administered to the organism, or of promoters

that are only active at certain stages of development or in certain tissues. In addition, conditional expression may involve inactivating a gene, for example, by FLP/FRT- or Cre-lox-mediated recombination.

The term "restriction fragment length polymorphism (RFLP) analysis," as used herein, refers to a method of determining whether an organism carries a specific nucleic acid sequence, for example, a specific alteration in a gene. This method may involve, for example, amplification of a nucleic acid from the organism, followed by cleavage of the nucleic acid with an enzyme, such as a restriction enzyme, and visualizing the products of the cleavage reaction. Furthermore, the cleavage products may be compared to control reactions.

As used herein, "alters proliferation" refers to any change in the proliferation of a cell. For example, this term can be used to describe an increase or a decrease in the rate of cell division. In addition, an alteration of proliferation may refer to a normally quiescent cell entering into the cell cycle or a normally dividing cell ceasing to enter into the cell cycle.

"Measuring protein levels," as used herein, includes any standard assay used in the art to either directly or indirectly determine protein levels. Such assays, for example may include the use of an antibody, Western analysis, Bradford assays, and spectrophotometric assays.

"Measuring nucleic acid levels," as used herein, includes any standard assay used in the art to either directly or indirectly determine nucleic acid levels. Such assays include, for example, hybridization analysis, gel electrophoresis, Northern blots, Southern blots, and spectrophotometric assays.

By a "substantially pure polypeptide" is meant a polypeptide (for example, a Sal2 polypeptide) that has been separated from components that naturally accompany it. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a Sal2 polypeptide. A substantially pure Sal2 polypeptide may be obtained, for example, by extraction from a natural source (for example, a

mammalian cell); by expression of a recombinant nucleic acid encoding a Sal2 polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

By "isolated DNA" is meant DNA that is free of the genes which, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

Advantages

The tumor host range selection procedure described herein has significant advantages over genetic screens and biochemical approaches used in the past to identify viral functions and to elucidate aspects of the interaction between virus and host. For example, previous studies using conditional lethal mutants of the polyoma viruses failed to uncover the large T antigen function involving interaction with mSal2 despite the fact that this interaction is essential for virus growth both *in vitro* (e.g., in tissue culture) and *in vivo* (e.g., in the mouse). In contrast to the directed search for host range mutants based on complementation with integrated viral genes (Benjamin, *Proc. Natl. Acad. Sci. U.S.A.* 67:394-399 (1970)), the 'tumor host range' selection procedure of the invention is an undirected search utilizing non-polyoma transformed or tumor derived cells. Selection of virus mutants is therefore unbiased except for the possibility of being conditional on the transformed state of the particular permissive host being used. Thus, the inventive strategy can lead to the identification of viral functions and cellular targets not revealed by conventional genetic screens or co-immunoprecipitation.

Furthermore, the methods of the invention also have a particular advantage over standard chemotherapy treatments, and the like, in that they are specific for cells with a proliferative disease. Therefore, one would expect this type of therapy to have fewer toxic side effects than the chemotherapeutic agents used today.

Brief Description of the Drawings

Fig. 1 shows photographs of the growth of wild-type polyoma virus and the TMD25 virus on host cells.

Fig. 2A shows the 20 bp sequence duplication responsible for the TMD25 mutation.

Fig. 2B shows the interaction of mSal2 clones with wild-type polyoma virus proteins and the TMD25 virus proteins in a yeast two-hybrid assay.

Fig. 2C shows deletion analysis of the TMD25 mutant.

Fig. 3A shows the regions of the *mSal2* gene used to develop antibodies.

Fig. 3B shows antibody detection of $p150^{sal2}$ as the *mSal2* gene product in a Western blot using protein from mouse and human cells.

Fig. 3C shows a Western blot of extracts from human 293 and U2OS cells that was first probed with an antiserum against the mSal2 carboxyl-terminus. The filter was then stripped and re-probed with an antibody against the mSal2 amino-terminus.

Fig. 4A shows the binding of mSal2 to wild-type polyoma virus but not to TMD25 large T protein *in vitro*.

Fig. 4B shows the binding of mSal2 and wild-type, but not TMD-25 mutant, large T protein in transfected 3T3 cells. These results are confirmed in BMK cells infected with wild-type polyoma virus and with TMD25 mutant virus.

Fig. 5A shows the failure of TMD25 to replicate in newborn mice.

Fig. 5B shows that TMD25 fails to replicate in BMK cells and that $p150^{sal2}$ represses viral origin replication.

Fig. 6 shows a Western blot of mSal2 expression in various mouse tissues.

Fig. 7 shows a Western blot of hSal2 expression in human ovarian tumors.

Fig. 8 shows expression of $p150^{sal2}$ in human 293 cells.

Fig. 9 shows immunostaining of $p150^{sal2}$ in human ovary tissue (A) and in ovarian tumors (B).

Fig. 10A shows that $p150^{sal2}$ suppresses growth of human ovarian tumor cells, which is indicated by a reduction in BrdU incorporation in $p150^{sal2}$ transfected cells.

Fig. 10B shows a colony reduction assay that indicates that cells transfected with $p150^{sal2}$ are less viable than control transfected cells.

Fig. 11 is an agarose gel showing that the 73S allele is lost in some ovarian tumors.

Detailed Description of the Invention

The present invention provides a method for identifying genes that play a role in cancer as well as methods for diagnosing and treating patients who have cancers involving these genes.

Identifying genes altered in cancerous cells

Host range selection of viruses

The present invention describes the use of tumor host range mutant viruses (T-HR mutants) that are capable of replicating in abnormally proliferating cells but not in normal cells. Therefore, these viruses are useful for identifying genes altered in abnormally proliferating cells. T-HR mutants generally have a mutation that causes a modification of function of the protein encoded by that gene. These mutations typically lie in the transforming genes of the DNA tumor viruses and are usually activators of cellular proto-oncogenes or inactivators of tumor suppressor genes. T-HR mutants may be isolated based on their ability to propagate (i.e. to replicate and disseminate) only in tumor cells that have a mutation in a cellular protein that is normally targeted by a viral transforming protein.

The methods of the invention have been applied to a 'tumor host range' selection procedure using the polyoma virus as a tool to search for new interactions of viral proteins, e.g., T antigens, with cellular proteins. The rationale behind this approach is based on the idea that genetic changes in tumor cells resulting in a modification of function of the cellular protein can provide the basis for a search to uncover new viral functions and interactions with cellular targets. In principle, 'Tumor host range' selection could reveal mutations in other functions, e.g., VP1, 2 or 3 involving interactions with receptors or the cellular machinery involved in virus uptake, uncoating or transport to the nucleus, or even in some aspect of virus assembly, or enhancer mutations that lead to alterations in enhancer function.

For example, alterations in yet unknown targets of viral genes might occur in spontaneous tumors or non-virally transformed cells. This suggests a rationale for isolating T-HR mutants based on modification of function in cancer cells. Mutants selected to grow in tumor cells, but not in normal cells, are useful for identifying new viral gene functions and their cellular targets. Targets identified in this way may include products of tumor suppressor genes or proto-oncogenes or any factor expressed in normal cells, which the virus must inactivate in order to propagate, but that is no longer expressed in tumor cells.

Identification of mSal2

The utility of the T-HR mutant based approach for identifying new genes involved in the susceptibility to proliferative diseases is shown by the identification of *mSal2*. The use of a T-HR mutant coupled with the power of the yeast two-hybrid screen resulted in the identification of a cellular target protein. Using T-HR mutants to identify cell cycle regulatory proteins is advantageous on two levels; first, in choosing an appropriate wild-type 'bait' corresponding to the region altered in the mutant, and second, in enabling a counterscreen where lack of interaction with the mutant is helpful in identifying cellular target(s) relevant to the mutant phenotype and possibly also to the transformed state of the

permissive host. One embodiment of the general protocol included as an aspect of the invention is outlined in Table 1 below.

Table 1. Tumor Host Range Mutants - Selection Procedure and Target Identification

I. Mutant Selection

1. Random mutagenesis of wild-type viral DNA
2. Amplification of the mutant virus by growth in tumor cells
3. Cloning by plaque isolation on tumor cells
4. Screening of plaque lysates for the absence of growth in normal cells
5. Molecular cloning and sequencing of the mutant viral DNA

II. Target Identification and Validation

6. Screening of a mouse embryo cDNA library in yeast with wild-type bait
7. Counterscreening positive clones for lack of interaction with mutant bait
8. Construction of complete cDNA expressing the target protein
9. Verification of viral protein-cellular target interactions *in vitro* and *in vivo* (e.g., T antigen-cellular protein interactions).

III. Identification of Risk Factors

10. Sequencing DNA derived from a tumor
11. Sequencing DNA derived from normal tissue of the same patient
12. Using the sequence information to establish whether the mutation is somatic or germline
13. Using this information in an epidemiological study to assess risk factors in a population

What follows is an illustration of the use of the methods of the invention to identify a new target of large T antigen, referred to as mSal2, using T-HR mutants of the polyomavirus. First, tumor host range selection identified a host range mutant of the polyomavirus that is able to grow in certain tumor or

transformed cells but not in normal cells. The mutant virus encodes an altered large T antigen protein and is defective in replication and tumor induction in newborn mice. Next, mSal2 was identified as a binding target of the polyoma virus large T antigen through a yeast two-hybrid screen. mSal2 shows no interaction with the mutant large T antigen. Specifically, the mutant virus fails to bind mSal2 and is unable to propagate or to induce most of the tumor types in the mouse that the wild-type virus typically induces.

The gene product $p150^{sal2}$ is expressed in a number of mouse and human tissues. It is found in nuclei of germinal epithelial cells from normal human ovary but is missing or altered in ovarian carcinomas derived from these cells (Table 3). Using an antibody to mSal2 that cross-reacts with the human protein, Sal2 was shown to be expressed as a protein of approximately 150 kDa in several normal murine and human tissues. Normal human ovarian epithelial cells show strong nuclear staining with the antibody. A majority of ovarian carcinomas derived from these cells show no detectible $p150^{sal2}$ by Western analysis and are negative by *in situ* immunochemistry. Some tumors display diffuse cytoplasmic, rather than nuclear, staining. (See Examples below.)

mSal2 is a zinc finger protein and a putative transcription factor that may have a role as a tumor suppressor. *mSal2* is homologous to the *Drosophila* homeotic gene *spalt* and to *sal* homologues identified in several vertebrate species (see below). The human homologue of the *Drosophila spalt* gene, *hSal2*, has been mapped adjacent to, or overlapping with, a chromosomal region associated with a loss of homozygosity in ovarian and other cancers.

The *spalt* or *sal* gene family of transcription factors is conserved in evolution from flies to man. First identified in *Drosophila*, *spalt* is a region-specific homeotic gene which functions in specifying anterior and posterior structures in the early embryo (Kuhnlein et al., *EMBO J* 13:168-179 (1994); Jurgens et al., *EMBO J* 7:189-196 (1988)) and also in later stages of organogenesis (Kuhnlein et al., *Mech. Dev.* 66:107-118 (1997); Barrio et al., *Dev. Biol.* 215:33-47 (1999)). *spalt*-related *sal* genes have been identified and studied in worms (Basson et al., *Genes Dev.* 10:1953-1965 (1996)), fish (Koster et al.,

Development 124:3147-3156 (1997)), frogs (Hollemann et al., *Mech. Dev.* 55:19-32 (1996); Onuma, *Biochem. Biophys. Res. Commun.* 264:151-156 (1999)), mice (Ott et al., *Mech. Dev.* 56:117-128 (1996); Kohlhase et al., *Nat. Genet.* 18:81-83 (2000)) and man (Kohlhase et al., *Genomics* 38:291-298 (1996); Kohlhase et al., *Genomics* 1:216-222 (1999); Kohlhase et al., *Cytogenet. Cell Genet.* 84:31-34 (1999)). In humans, a defect in the *hSal1* gene underlies the multiple developmental defects seen in Townes-Brocke syndrome (Kohlhase et al., *Nat. Genet.* 18:81-83 (1998)). Sal proteins contain multiple Zinc fingers, which frequently occur as C2H2 pairs with a conserved motif (Kuhnlein et al., *EMBO J* 13:168-179 (1994)). mSal2 has a structural arrangement typically seen in vertebrates with a single finger (C3H) near the amino-terminus and a cluster of three fingers (C2H2) considered essential for DNA binding in the middle portion of the protein (Pabo et al., *Annu. Rev. Biochem.* 61:1053-1095 (1992)). Like other Sal proteins, mSal2 has both glutamine-rich and proline- and alanine-rich sequences consistent with its transcriptional activator and repressor functions.

Although it has been shown in several species that Sal family transcription factors play important roles in embryonic development, downstream target genes have yet to be identified. Nevertheless, two important signaling pathways lying upstream of *sal* have been recognized. Regulation of *spalt* occurs in part through *dpp*, a member of the TGF-β family, which functions as a 'gradient morphogen' in the early *Drosophila* embryo (de Celis et al., *Nature* 381:421-424 (1996); Lecuit et al., *Nature* 381:387-393 (1996); Nellen et al., *Cell* 85:357-368 (1996)). In *Medaka*, *Sal1* expression occurs in response to *hh* (*hedgehog*) and is downregulated through PK-A (Koster et al., *Development* 124:3147-3156 (1997)). The TGF-β family of polypeptides has well known inhibitory effects on epithelial cell growth and survival. Disruptions in signaling pathways initiated by TGF-β are known to occur in some cancers (Kretzschmar et al., *Current Opinion in Genetics & Development* 8:103-111 (1998); Serra et al., *Nature Med.* 2:390-391 (1996)). In particular, mutations in *SMAD* genes, essential mediators of signaling via TGF-β receptors, have been linked to pancreatic, colorectal, and other cancers (Eppert et al., *Cell* 86:543-552 (1996);

Hahn et al., *Science* 271:350-353 (1996); Schutte et al., *Cancer Res.* 56:2527-2530 (1996)). Similarly, disruptions in signaling *via* '*hedgehog*' ligands and their '*patched*' receptors are important in development of basal cell carcinoma (Hahn et al., *Cell* 85:841-851 (1996); Johnson et al., *Science* 272:1668-1671 (1996); Oro et al., *Science* 276:817-821 (1997); Stone et al., *Nature* 384:129-134 (1996)).

Diagnosis

Diagnosis and Risk Assessment

In addition to helping identify genes that are altered in cancerous cells, target gene profiles can also be used to diagnose and/or stage various proliferative disorders and for diagnosing pre-symptomatic genetic lesions in normal tissues. The methods of the present invention can be used to diagnose cancerous cells in a patient by determining whether the cells of the patient can act as permissive hosts for the growth of a mutant virus, particularly a T-HR mutant. As described above, a permissive host for the growth of a mutant virus (e.g., a mutant virus that lacks a functioning transforming protein) has a mutation in a cellular gene that is the target for the wild-type viral protein corresponding to the mutant viral protein. This cellular mutation is believed to compensate for the modification of function in a particular gene in the T-HR mutant and contribute to the cancerous phenotype of the cell.

Once a target protein has been identified, tests for the lack of interaction of the cellular protein with the mutant viral protein are used to confirm the specificity of the interaction of the cellular protein with the wild-type (transforming) protein. A lack of interaction indicates that binding of the wild-type viral protein to the cellular protein is specific. Protein interaction can be verified by numerous methods know to those skilled in the art, including, for example, yeast two-hybrid assays, GST-pull down assays, co-immunoprecipitation, and Far-Western analysis. General guidance regarding these techniques can be found in standard laboratory manuals, such as Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York,

NY, (1994)), and Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, N.Y., (1989)). Once an interaction between the wild-type viral protein and the cellular protein is confirmed, the complete gene and gene product can readily be identified by those skilled in the art using, for example, the methods described below.

The present invention recognizes that the T-HR mutant selection procedures identified herein may identify mutant cellular genes, and their encoded protein products, e.g., cellular genes encoding cell cycle proteins, tumor suppressors, proto-oncogenes, transcriptional factors, regulators of apoptosis, etc., that have genetic lesions associated with a particular proliferative disorder. Those skilled in the art will appreciate that many proliferative disorders, such as cancers, correlate with a particular mutation or mutations in the DNA of a patient. By comparing the sequence for a particular gene in both normal and tumor tissue from the same patient, one can determine if the mutation is of somatic or germline origin. This information that may be used to screen a population as a whole for individuals that are at an increased risk of developing a particular type of proliferative disorder.

The present invention provides a method of identifying a genetic lesion in a cell by determining whether a cell can act as a permissive host for the growth of a particular T-HR mutant, such a T-HR mutant virus being capable of growing on a cell having a specific genetic lesion and not being capable of growth on a cell lacking this genetic lesion. This type of information may even be used to further characterize the cancer cell (e.g., to grade the stage to which the cancer has progressed).

In addition, the cellular gene that encodes a protein that is a target for a viral transforming protein may also be analyzed to determine whether there is a genetic lesion in the cellular gene. Such a genetic lesion may be associated with a particular cancer. As noted above, the present inventors describe a genetic lesion that may be associated with ovarian cancer has been identified in a *Sal2* gene. Specifically, this genetic lesion, corresponding to the substitution of a Cys for the Ser at position 73 in protein encoded by the *mSal2* gene of SEQ ID NO:4,

has been identified in DNA from blood samples from patients with ovarian cancer. Probes and primers based on this genetic lesion may be used as markers to detect the Ser73Cys change in samples from other patients.

A genetic lesion in a candidate gene may be identified in a biological sample obtained from a patient using a variety of methods available to those skilled in the art. Generally, these techniques involve PCR amplification of nucleic acid from the patient sample, followed by identification of the genetic lesion by either altered hybridization, aberrant electrophoretic gel migration, restriction fragment length polymorphism (RFLP) analysis, binding or cleavage mediated by mismatch binding proteins, or direct nucleic acid sequencing. Any of these techniques may be used to facilitate detection of a genetic lesion in a candidate gene, and each is well known in the art; examples of particular techniques are described, without limitation, in Orita et al. (*Proc. Natl. Acad. Sci. USA* 86:2766-2770, 1989) and Sheffield et al. (*Proc. Natl. Acad. Sci. USA* 86:232-236 (1989)). Furthermore, expression of the candidate gene in a biological sample (e.g., a biopsy) may be monitored by standard Northern blot analysis or may be aided by PCR (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994); *PCR Technology: Principles and Applications for DNA Amplification*, H.A. Ehrlich, Ed., Stockton Press, NY; Yap et al., *Nucl. Acids. Res.* 19:4294 (1991)).

Once a genetic lesion is identified using the methods of the invention (as is described above), the genetic lesion is analyzed for association with an increased risk of developing a proliferative disorder. In this respect, the present invention provides a method of detecting the presence of a genetic lesion in a human *Sal2* gene in a physiological sample, however the method is not limited to the *Sal2* gene, but rather can be applied to any gene that is associated with an increased risk for developing a proliferative disorder.

Furthermore, antibodies against a protein produced by the gene included in the genetic lesion, for example the Sal2 protein. Antibodies may be used to detect altered expression levels of the protein, including a lack of expression, or a change in its mobility on a gel, indicating a change in structure or size. In

addition, antibodies may be used for detecting an alteration in the expression pattern or the sub-cellular localization of the protein. Such antibodies include ones that recognize both the wild-type and mutant protein, as well as ones that are specific for either the wild-type or an altered form of the protein, for example, one encoded by a polymorphic *Sal2* gene. Monoclonal antibodies may be prepared using the Sal2 proteins described above and standard hybridoma technology (see, e.g., Kohler et al., *Nature* 256:495 (1975); Kohler et al., *Eur. J. Immunol.* 6:511 (1976); Kohler et al., *Eur. J. Immunol.* 6:292 (1976); Hammerling et al., In *Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, New York, NY (1981); Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)). Once produced, monoclonal antibodies are also tested for specific Sal2 protein recognition by Western blot or immunoprecipitation analysis (by the methods described in, for example, Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)).

Antibodies used in the methods of the invention may be produced using amino acid sequences that do not reside within highly conserved regions, and that appear likely to be antigenic, as analyzed by criteria such as those provided by the Peptide Structure Program (Genetics Computer Group Sequence Analysis Package, Program Manual for the GCG Package, Version 7, 1991) using the algorithm of Jameson and Wolf (*CABIOS* 4:181 (1988)). These fragments can be generated by standard techniques, e.g., by the PCR, and cloned into the pGEX expression vector (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)). GST fusion proteins are expressed in *E. coli* and purified using a glutathione agarose affinity matrix as described in Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY, (1994)).

To generate rabbit polyclonal antibodies, and to minimize the potential for obtaining antisera that is non-specific, or exhibits low-affinity binding, two or three fusions are generated for each protein, and each fusion is injected into at least two rabbits. Antisera are raised by injections in series, preferably including

at least three booster injections. These methods for antibody production and characterization are applicable to any other protein that is identified by the methods of the invention.

The antibody may be used in immunoassays to detect or monitor protein expression, e.g., Sal2 protein expression, in a biological sample. A polyclonal or monoclonal antibody (produced as described above) may be used in any standard immunoassay format (e.g., ELISA, Western blot, or RIA) to measure polypeptide levels. These levels may be compared to normal levels. Examples of immunoassays are described, e.g., in Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)). Immunohistochemical techniques may also be utilized for protein detection. For example, a tissue sample may be obtained from a patient, sectioned, and stained for the presence of Sal2 using an anti-Sal2 antibody and any standard detection system (e.g., one which includes a secondary antibody conjugated to horseradish peroxidase). General guidance regarding such techniques can be found in, e.g., Bancroft and Stevens (*Theory and Practice of Histological Techniques*, Churchill Livingstone (1982); and Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)).

Use of hSal2 as a Diagnostic Tool

As an example of the utility of this approach, the likelihood that *hSal2* functions as a tumor suppressor for ovarian cancer has been explored directly by screening a number of ovarian carcinomas for expression of $p150^{sal2}$ and for mutations in the gene. Approximately 80% of the tumors examined were negative or showed altered or reduced patterns of expression by Western analysis. Immunolocalization in frozen tissue sections showed strong staining in nuclei of epithelial cells on the surface of the normal ovary. In most instances, tumor cells showed a complete lack of staining. However, when staining was present in otherwise negative tumors, cytoplasmic rather than nuclear staining was seen in some areas.

Further evidence for *hSal2* function as a tumor suppressor comes from a limited screen for mutations in *hSal2*, which uncovered point mutations in four cases. In addition, cytogenetic approaches and sequencing efforts utilizing microsatellite markers have been used to map *hSal2* adjacent to, and possibly overlapping with, a chromosomal region associated with loss of homozygosity in ovarian (Bandera et al., *Cancer Res.* 57:513-515 (1997)) and other cancers, e.g., bladder cancer (Chang et al., *Cancer Res.* 55:3246-3249 (1995)). Such approaches may continue to be used to map *hSal2* more precisely.

The *mSal2* gene identified by the present invention may further be used to elucidate the cellular pathways of tumor suppression that regulate key cell cycle events. Alternatively, *mSal2* may be used to screen for potential tumors, e.g., lung tumors, brain tumors, stomach tumors, prostate tumors; ovarian tumors, tumors in SCID mice, as well as in knockout or transgenic animals, as discussed in detail below.

Treatment

In addition to providing a method for identifying genes altered in cancer cells and diagnosing patients who carry such mutation, the invention further provides a method of killing an abnormally proliferating cell using a tumor host range mutant virus.

For example, T-HR mutants may be used to specifically target and kill cancer cells in an organism. Since these viruses can only propagate in cells that carry a mutation in a cellular gene that the virus would normally have to activate, in the case of proto-oncogene, or inactivate, in the case of a tumor suppressor gene, in order to propagate, such a virus would be specific to abnormal cells. Therefore, T-HR mutants can be used to specifically eliminate cancer cells from a patient. For example, a T-HR mutant (i.e., a polyomavirus carrying an altered large T antigen causing it to be defective in replication and tumor induction) may be used to selectively kill human ovarian cancer cells that carry a genetic lesion in the *hSal2* gene, such as the Ser73Cys substitution described above.

However, one skilled in the art would realize that any number of genes, including ones involved in cell growth, cell cycle regulation, and apoptosis, may be altered in cancer cells. The methods of the invention are applicable to any alteration in a cancer cell that allows a T-HR mutant to grow. Therefore, any cancer that enables a T-HR mutant to propagate can be treated according to the methods of the invention disclosed herein.

The therapeutic T-HR mutant may be administered by any of a variety of routes known to those skilled in the art, such as, for example, intraperitoneal, subcutaneous, parenteral, intravenous, intramuscular, or subdermal injection. However, the T-HR mutant may also be administered as an aerosol, as well as orally, nasally, or topically. Standard concentrations used to administer a T-HR mutant include, for example, 10^2, 10^3, 10^4, 10^5, or 10^6 plaque forming units (pfu)/animal, in a pharmacologically acceptable carrier. Appropriate carriers or diluents, as well as what is essential for the preparation of a pharmaceutical composition are described, e.g., in *Remington's Pharmaceutical Sciences* (18^{th} edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA, a standard reference book in this field.

Formulations for parenteral administration may, for example, contain excipients, sterile water, or saline. For inhalation, formulations may contain excipients, such as lactose. Furthermore, aqueous solutions may be used, for example, for administration in the form of nasal drops, or as a gel for topical administration. The exact dosage used will depend on the severity of the condition (e.g., the size of the tumor), or the general health of the patient and the route of administration. The T-HR mutant may be administered once, or it may be repeatedly administered as part of a regular treatment regimen over a period of time.

Compounds that may be tested for an effect on proliferative diseases can be from natural as well as synthetic sources. Those skilled in the field or drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the methods of the invention. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-,

prokaryotic-, or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceanographics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries may be produced, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

Transgenic and Knockout Animals

The present invention provides transgenic and knockout animals that develop ovarian tumors and accurately recapitulate many of the features of the human ovarian tumor, an important contribution, since animal models of ovarian carcinoma are currently not available. Without limitation, particularly preferred transgenic or knockout animals are those in which the tumorigenic phenotype is fully penetrant, the rate of progression of the neoplasm is rapid, and/or the lifespan of the transgenic or knock-out animal is not shortened by a knockout- or transgene-related pathology in other organs. Of course, it will be appreciated that these traits are not required.

The generation of transgenic or knockout mice may provide a valuable tool for the investigation of human ovarian cancer by generating a mouse model for studying the disease, based on the description of the human *Sal2* gene provided above. Preferably, the *hSal2* gene is used to produce the transgenic mice or the *mSal2* gene is the target of the knockout. However, other *Sal2* genes

may also be used to produce transgenic mice provided that they are compatible with the mouse genome and that the protein encoded by this gene is able to carry out the function of the mSal2 protein.

Furthermore, a transgene, such as a mutant *Sal2* gene, may be conditionally expressed (e.g., in a tetracycline sensitive manner). For example, the promoter for the *Sal2* gene may contain a sequence that is regulated by tetracycline and expression of the Sal2 gene product ceases when tetracycline is administered to the mouse. In this example, a tetracycline-binding operator, tetO, is regulated by the addition of tetracycline, or an analog thereof, to the organism's water or diet. The tetO may be operably-linked to a coding region, for example a mutant *Sal2* gene. The system also may include a tetracycline transactivator (tTA), which contains a DNA binding domain that is capable of binding the tetO as well as a polypeptide capable of repressing transcription from the tetO (e.g., the tetracycline repressor (tetR)), and may be further coupled to a transcriptional activation domain (e.g., VP16). When the tTA binds to the tetO sequences, in the absence of tetracycline, transcription of the target gene is activated. However, binding of tetracycline to the tTA prevents activation. Thus, a gene operably-linked to a tetO is expressed in the absence of tetracycline and is repressed in its presence. The tetracycline regulatable system is well known to those skilled in the art and is described in, for example, WO 94/29442, WO 96/40892, WO 96/01313, and Yamamoto et al. (*Cell* 101:57-66 (2000).

In addition, the knockout organism may be a conditional knockout. For example, FRT sequences may be introduced into the organism so that they flank the gene of interest. Transient or continuous expression of the FLP protein may then be used to induce site-directed recombination, resulting in the excision of the gene of interest. The use of the FLP/FRT system is well established in the art and is described in, for example, U.S. Patent Number 5,527,695, and in Lyznik et al. (*Nucleic Acid Research* 24:3784-3789 (1996)).

Conditional knockout organisms may also be produced using the Cre-lox recombination system. Cre is an enzyme that excises DNA between two recognition sites termed loxP. The *cre* transgene may be under the control of an

inducible, developmentally regulated, tissue specific, or cell-type specific promoter. In the presence of Cre, the gene, for example a *Sal2* gene, flanked by loxP sites is excised, generating a knockout. This system is described, for example, in Kilby et al. (Trends in Genetics 9:413-421 (1993)).

Particularly preferred is a mouse model for ovarian cancer wherein the nucleic acid encoding a *Sal2* gene is expressed in the cells of the ovary of the transgenic mouse such that the transgenic mouse develops ovarian tumors. The mice preferably contain a large T antigen transgene, such as one expressing an appropriate (carboxyl-terminal) fragment of large T antigen under the control of an ovarian specific promoter, or have a knockout of the *mSal2* gene. In addition, ovarian cell lines from these mice may be established by methods standard in the art.

Transgenic animals may be made using standard techniques. For example, a gene encoding a cellular proto-oncogene, tumor suppressor gene, or other cellular protein, e.g., a cell cycle regulating protein, may be provided using endogenous control sequences or using constitutive, tissue-specific, or inducible regulatory sequences. Any tissue specific promoter may direct the expression of any Sal2 protein used in the invention, such as ovarian specific promoters, bladder specific promoters, and colon specific promoters. For example, knockout mutations may be engineered in the gene encoding the proto-oncogene or tumor suppressor gene and the mutated gene may be used to replace the wild-type *Sal2* gene.

Construction of transgenes can be accomplished using any suitable genetic engineering technique, such as those described in Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, N.Y., (1989)). Many techniques of transgene construction and of expression constructs for transfection or transformation in general are known and may be used for the disclosed constructs. Although the use of *hSal2* in the transgene constructs is used as an example, any other protein encoded by an oncogene may also be used.

One skilled in the art will appreciate that a promoter is chosen that directs expression of the oncogene in the tissue in which cancer is expected to develop. For example, as noted above, any promoter that regulates expression of *hSal2* in ovarian cancer cells can be used in the expression constructs of the present invention. Preferred ovarian promoters include, for example, promoters that are expressed in ovarian epithelial cells, such as, the polyoma virus promoter, the SPARK promoter, and the DOC-2 promoter. One skilled in the art would be aware that the modular nature of transcriptional regulatory elements and the absence of position-dependence of the function of some regulatory elements, such as enhancers, make modifications such as, for example, rearrangements, deletions of some elements or extraneous sequences, and insertion of heterologous (i.e., foreign) elements possible. Numerous techniques are available for dissecting the regulatory elements of genes to determine their location and function. Such information can be used to direct modification of the elements, if desired. It is preferred, however, that an intact region of the transcriptional regulatory elements of a gene is used. Once a suitable transgene construct has been made, any suitable technique for introducing this construct into embryonic cells can be used, an example of such a technique is provided in Example 9.

Animals suitable for transgenic experiments can be obtained from standard commercial sources such as Taconic (Germantown, NY). Many strains are suitable, but Swiss Webster (Taconic) female mice are preferred for embryo retrieval and transfer. B6D2F (Taconic) males can be used for mating and vasectomized Swiss Webster studs can be used to stimulate pseudopregnancy. In addition, vasectomized mice and rats are also publicly available from the above-mentioned suppliers. However, one skilled in the art would also know how to make a transgenic mouse or rat. An example of a protocol that can be used to produce a transgenic animal is provided in Example 9.

Use of Transgenic and Knockout Animals

The disclosed transgenic and knockout animals may be used as research tools to determine genetic and physiological features of a cancer, and for identifying compounds that can affect ovarian and other tumors. Knockout animals also include animals where the normal gene has been inactivated or removed and replaced with a mutant form of this gene, for example, a polymorphic allele. These animals can serve as a model system for assessing the risk of acquiring a proliferative disease that is associated with a particular mutation.

In general, the method of identifying markers associated with a proliferative disorder, such as ovarian tumors, involves comparing the presence, absence, or level of expression of genes, either at the RNA level or at the protein level, in tissue from a transgenic or knockout animal as described above, and in tissue from a matching non-transgenic or knockout animal. Standard techniques for detecting RNA expression, e.g., by Northern blotting, or protein expression, e.g., by Western blotting, are well known in the art. Differences between animals such as the presence, absence, or level of expression of a gene indicate that the expression of the gene is a marker associated with a proliferative disorder, such as ovarian tumors. The molecular markers, once identified, can be used to predict whether patients with carcinoma will have indolent or aggressive disease, and may be mediators of disease progression. Identification of such mediators would be useful since they are possible therapeutic targets. Identification of markers can take several forms.

One method by which molecular markers may be identified is by use of directed screens. Patterns of accumulation of a variety of molecules that may regulate growth can be surveyed using immunohistochemical methods. Screens directed at analyzing expression of specific genes or groups of molecules implicated in pathogenesis can be continued during the life of the transgenic or knockout animal. Expression can be monitored by immunohistochemistry as well as by protein and RNA blotting techniques. Mestastatic foci, once formed, can also be subjected to such comparative surveys.

Alternatively, molecular markers may be identified using genomic screens. For example, ovarian tissue can be recovered from young transgenic or knockout animals (e.g, that may have early stage carcinoma) and older transgenic or knockout animals (e.g., that may have advanced stage carcinoma), and compared with similar material recovered from age-matched normal littermate controls to catalog genes that are induced or repressed as disease is initiated, and as disease progresses to its final stages. These surveys will generally include cellular populations in the ovary.

This analysis can also be extended to include an assessment of the effects of various treatment paradigms (including the use of compounds identified as affecting ovarian tumors in the transgenic or knockout animals) on differential gene expression (DGE). The information derived from the surveys of DGE can ultimately be correlated with disease initiation and progression in the transgenic or knockout animals.

The following examples are meant to illustrate the invention and should not be construed as limiting.

Examples

Example 1: Isolation Of TMD-25 Using A 'Tumor Host Range' Selection

A procedure for isolating 'tumor host range' mutants (e.g., T-HR mutants) and identifying cellular targets is outlined below.

Identification of a Host Factor that interacts with T Antigens

1) Select Host Range Mutants
2) Identify Host Range Mutations
3) Identify Host Range Target and Validation
4) Biological Properties:
    (i) Viral DNA Replication
    (ii) Transformation
    (iii) Tumorigenicity

Permissive hosts were chosen based on a screen of mouse cell lines derived from non-polyoma-induced tumors or transformed cells using the following criteria: (i) susceptibility to lytic infection by wild-type polyoma virus, and (ii) ability to be used in standard plaque assays.

Among a number of qualifying cell lines, two were chosen: A6241, derived from a spontaneous mammary tumor in a C57BR mouse, and TCMK-1, a SV40-transformed baby mouse kidney cell line. Primary baby mouse kidney epithelial cells (BMK) were used throughout as the non-permissive host.

Randomly mutagenized virus was prepared by passage of a plasmid containing wild-type polyoma viral DNA through the error prone Mut D strain of *E. coli*, followed by excision of the viral genome and transfection into TCMK-1 cells. After several cycles of virus growth in the same cells, individual plaques were isolated using TCMK-1 cells. An aliquot of virus in each plaque suspension was inoculated into BMK cell cultures. Virus from plaques that induced no cytopathic effect (CPE) on BMK cells after 10-14 days was amplified using TCMK-1 cells. Mutant DNAs were cloned, reconstituted as virus by transfection of permissive cells, and confirmed to retain the desired host range. The frequency of mutants was approximately one in several thousand plaques tested. The T-HR mutant TMD-25 was isolated by this procedure.

Fig.1 shows the results of CPE tests comparing wild-type polyoma virus and TMD-25 growth in BMK, TCMK-1, and A6241 cells. Primary baby mouse kidney cells (BMK), SV40 Large T antigen transformed mouse kidney cells (TCMK), and spontaneous mouse mammary tumor cells (A6241) were mock-infected (Mock), or infected with 2-5 pfu of wild-type polyoma virus (PTA) or of T-HR mutant TMD25. The photographs were taken 14 days post infection and show the different cytopathic effects of viral growth.

TMD25 mutants grew poorly, if at all, on primary BMK cells, but could grow on transformed or tumor-derived cells, while wild-type polyoma virus grew well on all three cell-types. Extensive CPE developed in the TCMK-1 and A6241 cultures infected by the TMD25 mutant. Infectious mutant virus was produced in these cultures, although with somewhat slower kinetics and with

lower final yields compared to wild-type virus. In contrast, no discernible CPE was noted in mutant-infected BMK cultures, even after extended periods of incubation of up to three weeks. Growth of TMD-25 on the spontaneous tumor line A6241 rules out the possibility that its growth depends strictly on complementation by SV40 large T antigen, which is expressed in TCMK-1.

Example 2: Sequencing Of TMD-25 And Screening For Targets In Yeast

The mutation in TMD-25 responsible for its 'tumor host range' was localized to the carboxyl-terminal half of polyoma large T antigen as a result of studies using chimeric viruses constructed by ligating complementary DNA fragments from TMD-25 and wild-type virus. A combination of marker rescue and sequence analysis of this region revealed a twenty base pair duplication (circled) in TMD-25 encompassing the carboxyl-terminus of large T antigen. The resulting frameshift leads to replacement of the last 12 amino acids by 11 foreign residues (underlined) (SEQ ID NOS:9 to 12) (Fig. 2A).

It is possible that the carboxyl-terminal region of large T antigen is involved in binding to some cellular target as an essential step in virus growth and that the mutation in TMD-25 abolishes this interaction. As a first step toward identifying a possible cellular target, a cDNA library constructed from 9.5 to 10.5 day-old mouse embryos was screened in yeast two-hybrid assays, using the carboxyl-terminal portion of normal large T antigen (amino acids 335-782) as bait.

Twenty-two positive clones were analyzed. Nineteen of these clones were represented by nine independent but overlapping cDNA sequences that centered around a sixty-six amino acid region (amino acids 900-965) encompassing a zinc finger pair in the carboxyl-terminal region of the mSal2 protein cDNAs (Fig. 2B, Left Panel, and discussed below). The identified sequences showed strong homology to the human gene *hSal2*, which is related to *spalt* in *Drosophila*.

The positive mSal2 clones did not interact with the carboxyl-terminus of TMD25 large T antigen, as indicated by the growth (+) of yeast colonies on

histidine minus plates when using normal polyoma large T antigen as bait, but no growth using TMD25 large T antigen as bait (Fig. 2B, Right Panel), consistent with the notion that the host range defect of TMD-25 is based on its inability to bind this protein. All the His+ yeast colonies were also LacZ positive.

On continuous propagation in permissive cells, the TMD-25 mutant proved to be unstable, giving rise to wild-type virus revertants. To obtain a stable mutant and to further pinpoint the region of large T antigen essential for binding, (SEQ ID NOS:13 to 21), an analysis of the wild-type bait construct was carried out using mSal2 interaction in yeast as an assay (Fig. 2C). Truncation of the last six amino acids had no perceptible effect, but further truncations into the P-L-K sequence at positions 774-776 resulted in a loss of interaction. A deletion of these three amino acids in the context of an otherwise intact large T antigen was sufficient to prevent interaction with mSal2 and to recreate the host range phenotype shown in Fig. 1. The large T antigen deletion mutant 774-776 is hereafter referred to as TMD-25. The original defect of TMD25 is underlined, and the three amino acid region is framed in Fig. 2.

Example 3: Validation Of mSal2 As A Target Of Large T Antigen

A complete cDNA was obtained using RACE. The sequence was found to be identical to that reported recently for mSal2, with a Glu rather than a Lys residue at position 350. The genomic sequence indicates two alternate short 5' exons each encoding 24 amino acids and one unique 3' exon encoding 980 amino acids. The overall homology with hSal2 is 85% using the Blast 2 Sequence program. Eight Zinc fingers are apparent in exon 2. These zinc fingers are organized in four groups with the carboxyl-terminal pair presumed to be an essential part of the large T antigen interaction domain (Figs. 2B and 3A). Fig. 3A shows the corresponding gene region of the mSal2 protein fragments used to develop antibodies. The exons are boxed, with the zinc fingers represented as stripes.

Fig. 3B shows the antibody detection of *in* vitro translated full-length mSal2 and $p150^{sal2}$ in mouse and human cells. A polyclonal antibody was made

in rabbits against a GST fusion protein containing 131 amino acids from the carboxyl-terminal large T antigen interaction domain. Extracts of mouse 624 and human 293 cell lines probed with this antibody show a single protein species migrating at approximately 150 kDa (Fig. 3B, Right Panel). A monoclonal antibody against a 108 amino acid amino terminal fragment spanning exons 1 and 2 was isolated (Fig. 3A). This antibody also detected mSal2 as a 150 kDa *in vitro* translation product (Tr), as well as a protein present in normal mouse brain extracts (Br)(Fig. 3B). This gene product of mouse and human origin is referred to as $p150^{sal2}$. To confirm that the single band from the human cell extract is hSal2, extracts from two human cell lines first were probed with the polyclonal antibody made against the carboxyl-terminus of mSal2. The filter was then stripped and reprobed with the anti-mSal2 amino-terminus polyclonal antibody. The identical band was detected with each of the two antibodies in the human cell lysates (Fig. 3C).

In vitro pull down assays were carried out using a GST fusion of the large T antigen interaction domain of $p150^{sal2}$ and extracts of lytically infected or transfected cells (Fig. 4A). The filter was blotted with an anti-large T antigen antibody. Lanes "a" to "c" show pulldown assays using wild-type polyoma, lytic infected BMK cells: lane "a" shows input extract from normal (WT) Py infected BMK cells; lane "b" shows cell extract from lane "a" pulled down with GST alone; lane "c" shows cell extract from lane "a" pulled down with GST-mSal2 fusion protein. Lanes "d" to "h" show pulldown assays using cell extracts of 3T3 cells transfected with WT large T antigen or TMD25 large T antigen cDNA: lane "d" shows the input extract from 3T3 transfected with WT large T antigen cDNA; lane "e" shows the input extract from 3T3 transfected with TMD25 large T antigen cDNA; lane "f" shows the extract of WT large T antigen cDNA transfected 3T3 cells pulled down with GST alone; lane "g" shows the extract of WT large T antigen cDNA transfected 3T3 cells pulled down with GST-mSal2 fusion protein; and lane "h" shows the extract of TMD25 large T antigen cDNA transfected 3T3 cells pulled down with GST-mSal2 fusion protein. Normal large T antigen synthesized during infection of BMK efficiently binds the GST-mSal2

fragment (lanes a to c). Comparing extracts of 3T3 cells transfected with either wild-type, or TMD-25, large T antigen cDNAs only the wild-type shows binding (lanes d to g).

To confirm the large T-p150^{sal2} interaction *in vivo*, 3T3 cells were doubly transfected with a vector expressing full length GST-mSal2 and either wild-type, or TMD-25 mutant, large T antigen cDNAs (Fig. 4B Left Panel). Cell extracts were pulled down with glutathione beads. After electrophoresis and transfer, the filter was blotted with anti-large T antigen antibody to show the binding of wild-type or mutant large T antigen. The same filter was blotted again with a monoclonal antibody against mSal2 to show that the level of expression of GST-mSal2 is similar in both the wild-type large T antigen and the TMD25 large T antigen experiments. Each lane is labeled and the input equaled 3% of the extracts used in the co-precipitation assay. Complexes containing normal large T antigen were readily recovered, but no evidence of binding was seen with the mutant large T antigen.

A further experiment was done to confirm the interaction between the large T protein and p150^{sal2} during a lytic viral infection. An extract of wild-type virus-infected BMK cells was prepared 24 hours post-infection and incubated with polyclonal serum made against the amino-terminal mSal2 fragment. The anti-mSal2 immunoprecipitate was separated and blotted with an anti-T monoclonal antibody. A portion of the large T antigen present in the virus-infected cell extract clearly immunoprecipitated with mSal2, showing that these two proteins interact (Fig. 4B Right Panel). Polyoma large T and p150^{sal2} most likely interact directly through their carboxyl-terminal regions, although additional factors may be involved in mediating the binding.

Example 4: TMD-25 Is Defective In Virus Growth And Tumor Induction In The Newborn Mouse

Newborn mice were inoculated with either wild-type or TMD-25 mutant virus and followed for development of tumors. The ability of TMD-25 to replicate and spread in the newborn mouse was examined by whole mouse

section hybridization (Dubensky et al., *J. Virol.* 65:342-349 (1991). At ten days post inoculation the mutant showed no signs of replication and spread while the wild-type virus established a disseminated infection with extensive replication in many tissues (Fig. 5A).

Tests for virus replication were carried out on ten-day old animals by whole mouse section hybridization using a ^{35}S-labelled viral DNA probe (Fig. 5A). Newborn mice were inoculated subcutaneously with TMD25 or PTA (1 X 10^6 each) and sacrificed ten days later. Frozen sections were probed with ^{35}S labeled viral DNA with overnight exposure. Wild-type PTA showed strong replication in kidney, skin, and bones, while the TMD25 mutant showed no sign of viral replication in any of the organs. Table 2 shows a comparison of tumor induction profile between mSal2 binding mutant TMD25 and wild-type PTA viruses. Newborn mice were inoculated as described above, and sacrificed five months later. Pathological examinations were performed for tumor profile. Wild-type virus rapidly established a disseminated infection and induced a broad spectrum of tumors (Table 2). In contrast, TMD-25 failed to replicate and spread. The only tumors found in mutant-infected mice were subcutaneous fibrosarcomas and these developed only at the site of virus inoculation. Since TMD-25 is defective in replication but retains normal middle and small T functions, these findings are consistent with the expectation that the input mutant virus would be able to infect and transform cells locally but be unable to spread and induce a broad spectrum of tumors.

Direct tests of the mutant's transforming ability were carried out using standard assays with an established line of rat embryo fibroblasts (Dahl et al., *Mol. Cell Biol.* 16:2728-2735 (1996)). Transformation of these cells does not depend on virus replication, and middle T alone suffices for transformation (Raptis et al., *Mol. Cell Biol.* 5:2476-2485 (1985)). Mutant virus-infected cells gave rise to foci resembling those induced by wild-type virus; cells derived from one such focus were confirmed, by DNA sequencing, to carry the mutant viral genome. Using DNA transfection followed by measuring colony formation in soft agar, transforming efficiencies were found to be essentially identical for

wild-type and mutant viral DNAs – approximately 10-20 colonies/10^5 cell/μg viral DNA. The failure of TMD-25 to induce tumors at sites distant from the site of inoculation is therefore not due to any defect in transforming ability, but rather to its inability to replicate and establish a disseminated infection.

To investigate whether binding of $p150^{sal2}$ by large T antigen is necessary for viral DNA replication, low molecular weight DNA from BMK cells infected by wild-type or mutant virus was extracted and analyzed by Southern hybridization. The results show clearly that the mutant was unable to replicate its DNA in the non-permissive host (BMK) cells 36 hr post infection (Fig. 5B, Left Panel). BMK cells were infected with TMD25 and wild-type virus (Wt Py). Low molecular weight DNA was isolated at 0, 18, 36 hrs post infection (p.i.) for Southern blot with virus DNA probe. These results suggest that $p150^{sal2}$ can act, directly or indirectly, to inhibit viral DNA replication.

Furthermore, when over expressed in normal 3T3 cells, $p150^{sal2}$ inhibited wild-type viral DNA replication in a dose-dependent manner (Fig. 5B, Right Panel). Polyoma origin clone pUCori (Ori) and large T –expressing plasmid, (Wt LT cDNA), were cotransfected with increasing amount of plasmid expressing mSal2. Newly replicated DNA was detected with origin specific probe (top). The filter was striped and re-probed with LT and origin specific probe to show that similar amount of origin and LT DNA were present in each transfection. These results show that $p150^{sal2}$ imposes a block to viral DNA replication and that the block can be overcome by wild-type large T antigen.

Example 5: Expression Pattern Of $p150^{sal2}$ In The Mouse

Normal mouse tissues were extracted and tested for expression of $p150^{sal2}$ by Western blot (Fig. 6). Tissues from ten to twelve-day old mice were dissected and extracted in NP-40 lysis buffer. 200 μg of protein from various tissues were loaded onto each lane as labeled. The proteins were detected using a monoclonal antibody against the amino-terminus of mSal2. Tissue from brain, kidney, lung, bladder, and uterus clearly shows expression of the protein, while tissue from liver, skeletal muscle, spleen, salivary gland, and heart was either negative or low

in expression. These results are consistent with those reported earlier by Northern analysis. The finding that the kidney and lung are sites of strong expression is also consistent with the natural history of transmission of polyoma, which is thought to infect through the lung and amplify primarily in the kidney. Successful growth in these tissues would require the virus to be able to overcome any block to replication imposed by mSal2. TMD-25 fails to replicate its DNA in normal mouse cells, and overexpression of mSal2 blocks normal viral DNA replication.

Example 6: Expression Of hSal2 In Human Ovarian Tumors

The *hSal2* gene has been mapped to chromosome 14q12 but was not recognized initially as a tumor suppressor gene. It was subsequently shown by others that this region of 14q is associated with a loss of homozygosity in 49% of ovarian cancers (Bandera et al., *supra*) and about 25 % of bladder cancers (Chang et al., *supra*). These findings, along with the underlying rationale of 'tumor host range' selection, suggest the possibility that *sal2* may function as a tumor suppressor. To test this possibility more directly, a screen for $p150^{sal2}$ expression was carried out on extracts of ovarian carcinomas, the results of which are summarized in Fig. 7, a Western blot of human ovarian tumors. The expression level of $p150^{sal2}$ in 20 ovarian carcinomas was compared with that of normal ovarian epithelial cells (N) in two panels. Fifty micrograms of protein were loaded in each lane and blotted with polyclonal antibody against $p150^{sal2}$. Each ovarian carcinoma was labeled by its case number. Arrows indicate the normal position of p150. A polyclonal anti-p150 antibody made against the mouse protein clearly recognizes the human protein (Fig. 3B above). A band of the same apparent molecular weight is seen in extracts of normal human ovarian epithelial cells ('HOSE').

In situ staining with anti-p150 was carried out on frozen sections of normal ovary and several ovarian carcinomas, as well as in human 293 cells. Fig. 8 shows expression of $p150^{sal2}$ in human 293 cells. A polyclonal antibody, HM867, raised against mSal2 carboxyl-terminus, was used to detect human $p150^{sal2}$ in human 293 cells (lane +). As a negative control, the same protein

extract was blotted with HM867 antibody that had first been depleted by incubation with the same antigen used to raise it (lane -). As a further example of $p150^{sal2}$ expression, Fig. 9 shows immunostaining of $p150^{sal2}$ in the human ovary and in ovarian tumors. Fig. 9A shows immunostaining of normal human ovarian tissue with a polyclonal serum preadsorbed with mSal2 protein. In the left-hand panel, normal human ovarian tissue is stained with a polyclonal serum preadsorbed with $p150^{sal2}$. In the right-hand panel, normal ovarian tissue is stained with polyclonal serum against $p150^{sal2}$. Fig. 9B shows six ovarian carcinoma tissue samples that were stained for $p150^{sal2}$ (c thru h), where "T" stands for tumor cells and "S" stands for stromal cells. The insert in "h" shows cytoplasmic staining for $p150^{sal2}$. The nuclear staining of normal epithelial cells is readily apparent, but in the ovarian tumor cells the staining is reduced or cytoplasmic.

Example 7: A Point Mutation, S73C, In Human Sal2 Is Present In Some Ovarian Tumors.

DNAs from twenty-one ovarian carcinomas were digested and analysed by Southern hybridization using a probe of *hSal2* coding sequences. *hSal1* sequences were used as an unlinked internal control. No evidence of loss or gross rearrangement of the *hSal2* locus was seen in any of the tumors examined. However, deletions of 1kb or less would not have been detected. The absence of $p150^{sal2}$ expression in a majority of ovarian cancers may reflect mechanisms other than loss of the *hSal2* gene itself, such as silencing of expression through promoter methylation, alterations in an upstream regulatory factor, or factors leading to instability of the protein itself.

To test for small mutations, DNAs from four tumors were extracted and the entire *hSal2* coding regions sequenced on both strands. Two tumors from the panel shown in Fig. 7 that were positive for $p150^{sal2}$ expression and two that were negative were chosen. The two negative tumors 327 and 523 showed no changes when compared to the controls and all showed sequences identical to the published genomic sequence (Genbank AE000658 and AE000521; Boysen et al, *Genome Res.* 330:330-338 (1997)). The two $p150^{sal2}$–positive tumors each

showed a cysteine (TGT) substitution for serine (TCT) at position 73 (position 73 of SEQ ID NO:1), based on the first methionine in exon 1a (Kohlhase et al., *Mamm Genome* 11:64-69 (2000). The sequencing results showed only TGT in tumor 432 and a mixture of TGT and TCT in tumor 528. The serine codon TCT has been found at this position in all normal DNAs sequenced thus far (Kohlhase et al., *Genomics* 38:291-298 (1996); Boysen et al., *Genome Res.* 330:330-338 (1997)), indicating that '73S' is a frequent normal allele. To know whether the S73C substitution represents a somatic mutation or germ line polymorphism, normal DNA from case 432 was sequenced. The result showed only TGT at codon 73, indicating that the *hSal2* allele encoding cysteine represents a germ-line polymorphism in this individual. DNAs from six ovarian carcinoma cell lines were also sequenced and one showed the same S73C substitution as seen in case 432 and another a G744R substitution.

An example of the loss of the 73S allele is shown in Fig. 11. For this experiment, DNA was isolated from matched normal and ovarian tumor tissues. The 73S and 73C alleles were distinguished by PCR amplification and subsequent Mbo II digestion of a 318 bp product covering the region containing amino acid 73. In addition to a common Mbo II site (used to monitor the digestion status), this region contains another Mbo II site for the 73S allele, but not for the 73C allele (this is the discriminating Mbo II recognition site). Complete digestion of 73S allele by Mbo II produced three fragments (171 bp, 94 bp and 53 bp) while 73C allele produced two fragments (256 bp and 53 bp fragments-indicated by arrows). These fragments were resolved by electrophoresis on a 2% agarose gel. Although it is difficult to avoid the existence of normal tissue in the tumor used to isolate DNA, the intensity of the 73S bands (171 bp and 94 bp) is largely reduced indicating the loss of 73S allele (patient number 1). In this figure, "U" indicates undigested amplification product, "S" indicates a 73S homozygote control, "C" indicates a 73C homozygote control, and "S/C" indicates a 73S/C heterozygote control. The respective identification number of ovarian tumor patients is shown on top of their matched normal "N" and tumor "T" DNA.

Example 8: mSal2 Suppresses Growth of Ovarian Carcinoma Cells

To characterize the biological function of Sal2, the ovarian carcinoma cell line SKOV3 was transfected with an *mSal2* expression vector. SKOV3 cells were transfected with pcDNA-mSal2 (P150) or pcDNA3 vector (Mock), incubated in 0.5% serum for 48 hours, then in 15% serum and 100 μM BrdU for 20 hours. This cell line expresses little or no $p150^{sal2}$ as is indicated by Western analysis. Cells were examined by BrdU incorporation for DNA synthesis, for $p150^{sal2}$ expression, and for DAPI staining (Fig. 10A). The percent of cells in S-phase decreased from 57% in the control to 19% in cells expressing $p150^{sal2}$. In addition, 30-50% of cells expressing $p150^{sal2}$ appeared to be apoptotic as judged by DAPI staining compared to less than 10% of control cells. Arrows in frame 1 of Fig. 10A indicate a cell expressing $p150^{sal2}$ that is BrdU-negative. Arrows in frame 2 of Fig. 10A indicate an apoptotic cell expressing $p150^{sal2}$ with fragmented nuclear bodies as shown in the merged image. The bar graph in Fig. 10A shows the percentage of BrdU-positive cells in Mock and $p150^{sal2}$ expressing cells. In a colony reduction assay conducted over 14 days, a clear reduction in viable SKOV3 cells was seen in cells transfected with the expression vector, reflecting both growth suppressive, and apoptosis inducing activity of $p150^{sal2}$ (Fig. 10B). Similar efficiencies of transfection (approximately 20%) were confirmed by a co-transfected GFP expression plasmid.

Example 9: Experimental Procedures

Selection of tumor host-range mutants

The cell lines used as permissive hosts include TCMK-1 (Black et al., *Proc. Soc. Exper. Biol. Med.* 114:721-727 (1963)) purchased from ATCC) and A6241 (Lukacher et al., *J. Exp. Med.* 181:1683-1692 (1995); Velupillai et al., *J. Virology* 73: 10079-10085 (1999)). Primary baby mouse kidney cells (BMK) were used as the non-permissive host. The genome of polyoma virus strain PTA was digested at the single BamHI site and cloned into pBlueScript (Stratagene) to create PTAHI. PTAHI was amplified in the Mut D strain of *E. coli* (Schaaper et

al., *Proc. Natl. Acad. Sci. U.S.A.* 85:8126-8130 (1998)) to accumulate mutations randomly throughout the viral genome.

Yeast two-hybrid screening

The polyoma PTA large T antigen carboxyl-terminal fragment (amino acids 333-781) was cloned into pGBT9 (Clontech) to generate pGBT9ITC used as a "bait" to screen a 9.5 to 10.5 day-old whole mouse embryo cDNA library in pVP16 (Vojtek et al., *Cell* 75:205-214 (1993)). Transformation and selection were performed according to the recommendations from Clontech.

Generation of TMD25 with a minimum deletion

Large T antigen carboxyl-terminal deletions used in the yeast two-hybrid analysis were generated on pGBT9ITC using the Transformer site-directed *in vitro* mutagenesis kit (Promega) according to manufacturer's recommendations.

Cloning of full length mSal2 cDNA

A complete cDNA sequence for *mSal2* was obtained by RACE (Frohman) using Marathon cDNA amplification kit (Clontech) and RT-PCR products from BMK cells.

RFLP Test to Identify a Polymorphism in Sal2

Amino acid 73 of human $p150^{sal2}$ is polymorphic. This amino acid may be a serine encoded by the codon TCT (73S) or a cysteine encoded by the codon TGT (73C). The two alleles may be distinguished by PCR amplification of the genomic region encompassing the sequence encoding hSal2 amino acid 73 and digesting the PCR product using either the restriction enzyme Mob II or Ear I. These enzymes cut the DNA close to the codon encoding amino acid 73. The primers used to amplify the DNA prior to digestion with Mob II were, 5'-CTTGTTAATTAGAGCCTCGGTATACC-3' (SEQ ID NO:7) and 5'-GCACGGAGGACCCAGAATCTGG-3' (SEQ ID NO:8).

The PCR cycle used was 98°C for 2 minutes followed by 35 cycles of 94°C for 1 minute, 55°C for 1 minute, and 68°C for 1 minute. After the last PCR

cycle, the reaction was incubated at 72°C for 10 minutes. The PCR products were digested with Mob II in a solution containing 5 µl PCR mixture, 2 µl enzyme buffer (10 fold concentrated), 12 µl water, and 1 µl Mob II (5 units/µl). The restriction digest was performed at 37°C for two hours followed by heating the reaction to 70°C for twenty minutes prior to loading ten to twenty microliters of the mixture onto a 2% agarose gel. Five microliters of undigested PCR product were added to a control lane on the gel. The expected size of the uncut PCR product is 318 bp. The expected Mob II restriction fragments for the 73S allele are 171, 94, and 53 bp and the expected Mob II restriction fragments for the 73C allele are 265 and 53 bp. A mixture of the 73S and 73C alleles would be expected to yield fragments of 265, 171, 94, and 53 bp. The 53 bp fragment is common to both alleles and may be used to monitor the digestion status in order to distinguish between heterozygotes and an incomplete digestion.

In vitro GST pull-down assay

Full-length polyoma normal large T antigen cDNA and TMD25 large T antigen cDNA were cloned into pcDNA3 to create CMVLT and CMVTMDLT respectively. The mSal2 fragment (amino acids 841-971), containing the last zinc finger pair, was cloned into pGEX4T1 (Pharmacia) to generate GST-mSal2 fusion protein in *E. coli*. The fusion protein was bound to glutathione-Sepharose 4B beads (purchased from Pharmacia) according to the manufacturer's instructions. For the association of GST-mSal2 fusion with large T antigen, BMK cells infected by PTA, or 3T3 cells transfected with wild-type or TMD25 large T antigen expression constructs CMVLT or CMVTMDLT, were extracted with NP-40 lysis buffer (pH 7.9) (Benjamin et al., *Proc. Natl. Acad. Sci. U.S.A.* 67:394-399 (1970)). 500µl of cell lysate were incubated with 50µl of 50% GST-Sal2 or GST beads for 2 hours. After washing four times with PBS, the bound protein was subjected to Western blot analysis using monoclonal antibody F4, which recognizes T antigens (Dahl et al., *Mol. Cell. Biol.* 16:2728-2735 (1996)).

In vivo GST pull-down assay

The full-length mSal2 coding region was cloned into a eukaryotic GST fusion vector, pEBG (Luo et al., *J. Biol. Chem.* 270:23681-23687 (1995)) to generate the construct pEBGSAL. NIH 3T3 cells were co-transfected with pEBGSAL and CMVLT or CMVTNDLT in a ratio of 1 to 1 using Lipofect2000 (Gibco/BRL) according to the manufacturer's protocol. The cells were harvested 48 hours post transfection. The lysate was centrifuged at 3,000 rpm and the supernatant was incubated with 50-100 µl glutathione-Sepharose 4B beads for 2 hours. The beads were washed four times with PBS containing 0.01% NP-40 and the bound proteins were immunoblotted with the F4 antibody and an antibody against $p150^{sal2}$ (Dahl et al., *supra* (1996)).

In vivo Co-immunoprecipitation of mSal2 and Polyoma Large T

Fifty microliters of 50% protein A beads (Pharmacia) were incubated with purified rabbit polyclonal anti-amino-terminal mSal2 antibody or normal rabbit IgG in 1 ml NP-40 lysis buffer at 4°C for 2 hours, followed by washing the beads four times with PBS. Two milligrams of total protein, made from BMK cells infected with wild-type virus, were extracted 24 hours post infection and incubated with either the anti-mSal2 or normal lgG beads in NP-40 lysis buffer containing 1% BSA for 2 hours at 4°C. After the incubation, the beads were washed four times with 0.1% Tween-20 in PBS and the proteins were separated by SDS-PAGE. Polyoma large T and mSal2 were detected using anti-T and anti-mSal2 monoclonal antibodies.

Viral DNA Replication Assays

Plasmid pUCori and the polyoma origin replication assay are described in Gjorup et al. (*Proc. Natl. Acad. Sci.* USA 91:12125-12129 (1994)). Cells were grown on 6 well plates and infected with virus or transfected with DNA. Low molecular weight DNA was isolated as described by Hirt (*J. Mol. Biol.* 26:365-369 (1967)). After purification, the DNAs were resuspended in 80 µl of water. One to five micrograms of DNA were subjected to restriction digestion. For

virus infection experiments, the viral genome was first linearized with Eco RI. For transfection experiments, pUCori and CMVLT were first digested with Dpn I and Hind III. The newly synthesized pUCori DNA is Dpn I resistant because of the lack of methylation in eukaryotic cells and the input plasmid DNA is sensitive to Dpn I digestion because of the *E. coli* methylation of the recognition site. The DNA fragment was resolved on a 1% agarose gel for Southern analysis using origin specific and LT specific probes.

Western blots for detection of $p150^{sal2}$

Tissue extracts were prepared from C3H/BiDa mice by homogenization in NP-40 lysis buffer (pH 7.9) and centrifugation at 8,000 rpm. Fifty micrograms of protein (Bio-Rad Assay) from each sample was separated by SDS-PAGE and blotted on nitrocellulose membranes. A monoclonal antibody against mSal2 was used to detect $p150^{sal2}$.

Stripping Western Filters for Reprobing

After first antibody probing, the used filter is incubated in stripping solution (50 mM Tris-Cl, pH 6.8, 2% SDS and 100 mM β-mercaptoethanol) for 30 minutes at 60°C. The filter is washed twice in PBS and tested for the absence of the previously used antibody by development and exposure to an X-ray film. This procedure ensures that the filter can be used again in subsequent Western analyses.

Analysis of ovarian carcinomas

Surgical samples of human ovarian tissue were obtained under a protocol approved by the Human Subjects Committee of the Brigham and Women's Hospital. Ovarian tumor tissues were pulverized in liquid nitrogen and lysed in a buffer (1% Triton X-100, 21 μg/ml aprotinin, 1 mM phenylmethylsulfonyl fluoride, 0.5 μg/ml leupeptin, 4.9 mM $MgCl_2$, and 1mM vanadate in PBS). The MicroBAC Protein Assay Kit (Pierce) was used for protein quantitation.

Twenty-five micrograms of protein from each sample were separated on an SDS-polyacrylamide gel and blotted on nitrocellulose membranes. A rabbit polyclonal antiserum that cross-reacts with hSal2 was used to detect $p150^{sal2}$. Specifically, this antiserum was raised against a GST-mouse $p150^{sal2}$ fusion protein that was first purified using Affinity Pak Immobilized Protein A (Pierce) according to manufacturer's instructions followed by an incubation with GST saturated glutathione beads (Pharmacia) in PBS for 30 minutes to eliminate antibodies against GST. As a negative control, the purified antibody was preadsorbed with the GST- $p150^{sal2}$ fusion protein.

Frozen sections of normal or tumor samples were fixed in Neutral Formalin for 10 minutes and permeabilized in cold ethanol/acidic acid (3:1) for 15 min. After washing four times in PBS for 10 minutes each, the sections were antibody stained and processed using Vectastain ABC kit (Vector Laboratories) following the manufacturer's instructions.

DNAs were extracted from human ovarian carcinomas and from primary cultures of ovarian epithelial cells obtained by scraping the surface of normal ovarian tissue. DNA from normal human foreskin was used as a control. The coding region with the 0.4 kb intron of *hSal2* was amplified using the primer pair (5'-CCACAACCATGGCGAATCCGAG-3') (SEQ ID NO:5) and (5'-GGTGATGGAAGGCGAACAGCCAGG-3') (SEQ ID NO:6). Long range PCR was performed (98°C 4 min, then 94°C 1 min, 60°C 1 min, 68°C 4 min, for 35 cycles) and sequencing was carried out using the High Throughput Core of the Dana Farber-Harvard Cancer Center. The coding region was sequenced twice and additional sequencing of both strands was performed for regions with suspected mutations. The resulting sequence was compared with the published *hSal2* cDNA sequence and genomic sequence.

BrdU Incorporation

SKOV3 cells were transfected with pcDNA-*mSal* and the pcDNA 3 vector using BRL Lipofectamine 2000 according to the manufacturer's recommendations. Five to seven hours post transfection the cells were fed with

0.5% calf serum. After 48 hours, the cells were incubated with a medium containing 15% calf serum with 100 mM BrdU for 20 to 24 hours. A monoclonal antibody against BrdU (Amersham) was used to detect the incorporation. The cells were fixed, permeabilised and stained according to Amersham's recommendations except that a purified rabbit polyclonal antibody against the mSal2 carboxyl-terminus was mixed with the BrdU antibody for the detection of both BrdU incorporation and $p150^{sal2}$ expression. Secondary antibodies (anti-mouse Rhodamine and anti-rabbit Oregon Green) were also mixed. Cells were examined under fluorescence microscopy in order to identify BrdU and $p150^{sal2}$ positive cells.

Colony Reduction Assay

SKOV3 cells were transfected with a pcDNA-*mSal* or a pcDNA3 vector in a 6 well plates using 2 μg of DNA each. To monitor the transfection efficiency, 0.5 μg of pEGFPN1 (Clontech) was added to the test DNA in a separate tube. Transfection was performed according to GIBCO/BRL's recommendations using LIPOFECTAMIN 2000. Twenty-four hours after the transfection, the cells were re-seeded in 10 cm plates with medium containing 600 μg/ml G418 (GIBCO/BRL) and 10% calf serum. The EGFP expression was also monitored at this time. The G418 containing medium (neomycin medium) was changed every 3 to 4 days until mock-transfected cells had died and neomycin resistant colonies became apparent.

Preparation of DNA for microinjection

As but one example, DNA clones for microinjection are prepared by cleaving the DNA with enzymes appropriate for removing the bacterial plasmid sequences and subjecting the DNA fragments to electrophoresis on 1% agarose gels in TBE buffer (Sambrook et al. (1989)). The DNA bands are visualized by staining with ethidium bromide and the band containing the desired DNA sequences is excised. The excised band is then placed in dialysis bags containing 0.3 M sodium acetate pH 7.0. The DNA is electroeluted into the dialysis bags,

extracted with phenol/chloroform (1:1), and precipitated by the addition of two volumes of ethanol. The DNA is then redissolved in 1 ml of low salt buffer (0.2 M NaCl, 20 mM Tris, pH 7.4, and 1 mM EDTA) and purified on an Elutip-D.TM (Schleicher and Schuell) column. The column is first primed with 3 ml of high salt buffer (1M NaCl, 20 mM Tris, pH 7.4, and 1 mM EDTA) followed by washing with 5 ml of low salt buffer. The DNA solutions are passed through the column three times to bind the DNA to the column matrix. After one wash with 3 mls of low salt buffer, the DNA is eluted with 0.4 ml of high salt buffer and precipitated by the addition of two volumes of ethanol. DNA concentrations are measured by absorption at 260 nm in a UV spectrophotometer. For microinjection, DNA concentrations are adjusted to 5 μg/ml in 5 mM Tris, pH 7.4 and 0.1 mM EDTA. Other methods for purification of DNA for microinjection are also described in Hogan et al. (*Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1986)); in Palmiter et al. (*Nature* 300:611 (1982)); in the Qiagenologist, *Application Protocols*, 3rd edition, published by Qiagen, Inc., Chatsworth, Calif.; and in Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989)). The procedures for manipulation of the rodent embryo and for microinjection of DNA are described in detail in Hogan et al. (*Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1986)), the teachings of which are incorporated herein.

Animal experiments

Whole mouse section hybridizations (Dubensky et al., *J. Virol.* 68:342-349 (1991)) and tumor profiles (Dawe et al., *Am. J. Pathol.* 127:243-261 (1987)) were performed as described in these publications.

Production of transgenic mice and rats

The following is but one preferred means of producing transgenic mice. This general protocol may be modified by those skilled in the art.

Female mice six weeks of age are induced to superovulate with a 5 IU injection (0.1 cc, IP) of pregnant mare serum gonadotropin (PMSG; Sigma) followed 48 hours later by a 5 IU injection (0.1 cc, IP) of human chorionic gonadotropin (hCG, Sigma). Females are placed together with males immediately after hCG injection. Twenty-one hours after hCG injection, the mated females are sacrificed by CO_2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts and placed in Dulbecco's phosphate buffered saline with 0.5% bovine serum albumin (BSA, Sigma). Surrounding cumulus cells are removed with hyaluronidase (1 mg/ml). Pronuclear embryos are then washed and placed in Earle's balanced salt solution containing 0.5% BSA (EBSS) in a 37.5°C incubator with humidified atmosphere at 5% CO_2, 95% air until the time of injection. Embryos can be implanted at the two-cell stage.

Randomly cycling adult female mice are paired with vasectomized males. Swiss Webster or other comparable strains can be used for this purpose. Recipient females are mated at the same time as donor females. At the time of embryo transfer, the recipient females are anesthetized with an intraperitoneal injection of 0.015 ml of 2.5% avertin per gram of body weight. The oviducts are exposed by a single midline dorsal incision. An incision is then made through the body wall directly over the oviduct. The ovarian bursa is then torn with watchmakers forceps. Embryos to be transferred are placed in DPBS (Dulbecco's phosphate buffered saline) and in the tip of a transfer pipet (about 10 to 12 embryos). The pipet tip is inserted into the infundibulum and the embryos are transferred. After the transferring the embryos, the incision is closed by two sutures.

The preferred procedure for generating transgenic rats is similar to that described above for mice (Hammer et al., *Cell* 63:1099-112 (1990). For example, thirty-day old female rats are given a subcutaneous injection of 20 IU of PMSG (0.1 cc) and 48 hours later each female placed with a proven, fertile male. At the same time, 40-80 day old females are placed in cages with vasectomized males. These will provide the foster mothers for embryo transfer.

The next morning females are checked for vaginal plugs. Females who have mated with vasectomized males are held aside until the time of transfer. Donor females that have mated are sacrificed (CO_2 asphyxiation) and their oviducts removed, placed in DPBA (Dulbecco's phosphate buffered saline) with 0.5% BSA and the embryos collected. Cumulus cells surrounding the embryos are removed with hyaluronidase (1 mg/ml). The embryos are then washed and placed in EBSs (Earle's balanced salt solution) containing 0.5% BSA in a 37.5°C incubator until the time of microinjection.

Once the embryos are injected, the live embryos are moved to DPBS for transfer into foster mothers. The foster mothers are anesthetized with ketamine (40 mg/kg, IP) and xulazine (5 mg/kg, IP). A dorsal midline incision is made through the skin and the ovary and oviduct are exposed by an incision through the muscle layer directly over the ovary. The ovarian bursa is torn, the embryos are picked up into the transfer pipet, and the tip of the transfer pipet is inserted into the infundibulum. Approximately 10 to 12 embryos are transferred into each rat oviduct through the infundibulum. The incision is then closed with sutures, and the foster mothers are housed singly.

Generation of Knockout Mice

The following is but one example for the generation of a knockout mouse and the protocol may be readily adapted or modified by those skilled in the art.

Embryonic stem cells (ES), for example, 10^7 AB1 cells, may be electroporated with 25 μg targeting construct in 0.9 ml PBS using a Bio-Rad Gene Pulser (500 μF, 230 V). The cells may then be plated on one or two 10-cm plates containing a monolayer of irradiated STO feeder cells. Twenty-four hours later, they may be subjected to G418 selection (350 μg/ml, Gibco) for 9 days. Resistant clones may then be analyzed by Southern blotting after *Hind* III digestion, using a probe specific to the targeting construct. Positive clones are expanded and injected into C57BL/6 blastocysts. Male chimeras may be back-crossed to C57BL/6 females. Heterozygotes may be identified by Southern blotting and intercrossed to generate homozygotes.

In addition, knockout mice may also be generated by site-specific recombination methods using, for example, the FLP/FRT system or the Cre-lox system. These systems are described in the specification as well as in, for example, U.S. Patent Number 5,527,695, Lyznik et al. (*Nucleic Acid Research* 24:3784-3789 (1996), and Kilby et al. (Trends in Genetics 9:413-421 (1993)).

The targeting construct used in making the knockout animal may result in the disruption of the gene of interest, e.g., by insertion of a heterologous sequence containing stop codons, or the construct may be used to replace the wild-type gene with an altered form of the same gene, e.g., a mutant *Sal2* gene. In addition, the targeting construct may contain a sequence that allows for conditional expression of the gene of interest. For example, a sequence may be inserted into the gene of interest that results in the protein not being expressed in the presence of tetracycline. Such conditional expression of a gene is described in, for example, WO 94/29442, WO 96/40892, WO 96/01313, and Yamamoto et al. (*Cell* 101:57-66 (2000).

Table 2. Tumor profiles of mutant TMD-25 and wild-type PTA virus

	TMD-25	PTA[1]
Fraction of mice with tumors	7/7	32/32
Mean age at necropsy	202d.	82d.
Epithelial tumors:		
Hair follicle	0/7	32/32
Thymus	0/7	29/32
Mammary gland	0/7	16/32
Salivary gland	0/7	23/32
Mesenchymal tumors:		
Fibrosarcomas	7/7[2]	1/32
Renal medulla	0/0	7/32
Bone	0/0	6/32

1. Data on PTA is taken from Dawe et al. (*Am. J. Pathol.* 127:243-261, 1987).

2. Subcutaneous fibrosarcomas were found only at the site of virus inoculation.

Table 3. Summary of $p150^{sal2}$ expression in human ovarian carcinomas

$p150^{sal2}$ Status	Number of Cases	Percent
7		
Positive	6	30
Negative	10	50
Altered*	4	20

* Refers to the apparent size of the Sal2 protein, which is different from that of normal ovarian epithelial cells.

Other Embodiments

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth.

All references cited herein are hereby incorporated by reference.

We claim:

CLAIMS

1. A method of identifying a cellular protein involved in the susceptibility to proliferative disease, said method comprising the steps of:

a) infecting a normal cell and an abnormally proliferating cell with a collection of uncharacterized mutant viruses;

b) identifying a mutant virus from the collection that can grow in said abnormally proliferating cell and can not grow in said normal cell; and

c) identifying the mutated viral gene or mutated protein in said virus, which allows said virus to grow on said abnormally proliferating cell; and

d) screening to identify the cellular protein which interacts with the wild-type viral protein, but not said mutated viral protein.

2. The method of claim 1, wherein said abnormally proliferating cell is uncharacterized.

3. The method of claim 1, further comprising identifying a cellular protein that can interact with a wild-type viral protein that corresponds to said mutant viral protein, wherein said cellular protein is not a retinoblastoma tumor suppressor protein.

4. The method of claim 3, wherein the step of identifying said cellular protein comprises using an assay that detects protein-protein interactions.

5. The method of claim 4, wherein said assay is a GST-pulldown assay.

6. The method of claim 3, further comprising isolating a gene encoding said cellular protein.

7. The method of claim 1, wherein said virus has a mammalian host range.

8. The method of claim 7, wherein said mammal is a human.

9. The method of claim 1, wherein said virus is selected from the group consisting of simian virus 40 virus, human polyoma virus, parnovirus, papilloma virus, herpes virus, and primate adenoviruses.

10. The method of claim 1, wherein said cellular protein is a tumor suppressor protein.

11. The method of claim 1, wherein said cellular protein is a proto-oncogene product.

12. A tumor host range virus isolated using the method of claim 1.

13. A method of determining the presence or absence of an alteration in the genetic material of a cell, said method comprising determining whether a cell can act as a permissive host for the propagation of a characterized T-HR mutant, said T-HR mutant being capable of propagating in an abnormally proliferating cell and not being capable of propagating in a normal cell, wherein said characterized T-HR mutant is unable to propagate in a cell carrying a mutation in the retinoblastoma or p53 gene.

14. The method of claim 13, wherein the presence of said genetic alteration is indicative of an organism carrying this genetic alteration being at an increased risk of developing a proliferative disease.

15. The method of claim 13, wherein said alteration in the genetic material is in a tumor suppressor gene.

16. The method of claim 13, wherein said alteration in the genetic material is in a proto-oncogene.

17. The method of claim 13, wherein said characterized T-HR mutant has been characterized as being complemented by a mutation in a specific tumor suppressor gene or proto-oncogene, wherein said tumor suppressor or proto-oncogene are not the retinoblastoma or p53 gene.

18. The method of claim 13, wherein said cell is a cell from a mammal.

19. The method of claim 18, wherein said mammal is a human.

20. A method of killing an abnormally proliferating cell comprising the steps of:

(i) contacting an abnormally proliferating cell with a T-HR mutant; and

(ii) allowing said T-HR mutant to lyse said cell.

21. The method of claim 20, wherein said abnormally proliferating cell is a mammalian cell.

22. The method of claim 21, wherein said mammalian cell is a human cell.

23. The method of claim 20, wherein said abnormally proliferating cell is in a mammal with a proliferative disorder.

24. The method of claim 23, wherein said mammal is a human.

25. The method of claim 20, wherein said T-HR mutant is administered in a pharmaceutically acceptable carrier.

26. The method of claim 20, wherein said T-HR mutant is administered by a method selected from the group consisting of parenteral, intravenous, intraperitoneal, intramuscular, subcutaneous, and subdermal injection.

27. The method of claim 20, wherein said T-HR mutant is administered by a method selected from the group consisting of orally, nasally, topically, and as an aerosol.

28. The method of claim 20, wherein said TH-R mutant is selected from the group consisting of, simian virus 40, human polyoma virus, herpes virus, primate adenoviruses, parnovirus, and papilloma virus.

29. The method of claim 20, wherein said T-HR mutant is specific for a cell carrying a *Sal2* mutation.

30. The method of claim 29, wherein said T-HR mutant is the TMD-25 T-HR mutant virus.

31. A method of identifying a mammal having or at increased risk of acquiring a proliferative disease, said method comprising the step of determining whether there is a proliferative disease-associated alteration in a *Sal2* nucleic acid of said mammal.

32. The method of claim 31, wherein said method is for identifying a mammal having a proliferative disease.

33. The method of claim 31, wherein said method is for identifying a mammal at increased risk of acquiring a proliferative disease.

34. The method of claim 31, wherein said mammal is a human.

35. The method of claim 34, wherein said proliferative disease-associated alteration comprises the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

36. The method of claim 31, wherein said determining is done by polymerase chain reaction (PCR) amplification, single nucleotide polymorphism (SNP) determination, restriction fragment length polymorphism (RFLP) analysis, hybridization analysis, or mismatch detection analysis.

37. The method of claim 31, wherein said method comprises the steps of:

(i) contacting a first nucleic acid probe which is specific for binding to

said human *Sal2* nucleic acid containing said alteration with a nucleic acid from a cell from said mammal under conditions which allow said first nucleic acid probe to anneal to complementary sequences in said cell; and

(ii) detecting duplex formation between said first nucleic acid probe and said complementary sequences.

38. The method of claim 37, wherein said first nucleic acid probe is derived from the human *Sal2* nucleic acid containing a proliferative disease-associated alteration.

39. The method of claim 37, further comprising a second nucleic acid probe, wherein said first and second nucleic acid probes are PCR primers, and wherein said human *Sal2* nucleic acid or a fragment thereof is amplified using PCR between steps (i) and (ii).

40. The method of claim 37, wherein said cell is from a physiological sample containing abnormally proliferating tissue.

41. The method of claim 37, wherein said cell is from a physiological sample of normal tissue.

42. The method of claim 37, wherein said alteration comprises the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

43. A method of identifying a mammal having or at increased risk of acquiring a proliferative disease, said method comprising the step of determining whether there is a proliferative disease-associated alteration in a Sal2 protein of said mammal.

44. The method of claim 43, wherein said method is for identifying a mammal having a proliferative disease.

45. The method of claim 43, wherein said method is for identifying a mammal at increased risk of acquiring a proliferative disease.

46. The method of claim 43, wherein said mammal is a human.

47. The method of claim 43, wherein said method comprises the use of an antibody specific for a human Sal2 protein.

48. The method of claim 47, wherein said antibody comprises an antibody specific for a proliferative disease-associated mutant Sal2 protein.

49. A knockout mouse comprising a knockout mutation in a genomic *mSal2* gene.

50. The knockout mouse of claim 49, wherein said mouse further comprises a nucleic acid construct including a mutant *Sal2* gene.

51. The knockout mouse of claim 40, wherein said mutant *Sal2* gene is conditionally expressed.

52. The knockout mouse of claim 40, wherein said mutant *Sal2* gene encodes a protein that contains a substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

53. A transgenic mouse whose genome comprises a nucleic acid construct
including a *Sal2* nucleic acid, which is operably linked to transcriptional regulatory elements and encodes a Sal2 protein.

54. The transgenic mouse of claim 53, wherein said Sal2 protein is mutant.

55. The transgenic mouse of claim 53, wherein said transcriptional regulatory elements include a promoter that is a tissue-specific promoter.

56. The transgenic mouse of claim 55, wherein said nucleic acid is expressed such that the protein is produced at detectable levels in cells selected from the group consisting of ovarian, bladder, and colon cells.

57. The transgenic mouse of claim 55, wherein said transcriptional regulatory elements include a promoter that is an ovary-specific promoter.

58. The transgenic mouse of claim 53, wherein said Sal2 nucleic acid is a human Sal2 nucleic acid.

59. The transgenic mouse of claim 54, wherein said mouse develops ovarian tumors.

60. The transgenic mouse of claim 59, wherein said ovarian tumors metastasize.

61. A cell line derived from cells isolated from said transgenic mouse of claim 53.

62. A method of identifying a compound which alters cell proliferation, said method comprising:

a) contacting a first cell with a test compound, and

b) measuring whether said test compound alters proliferation in said first cell, relative to a second cell not contacted with said test compound, wherein said first and second cells have a proliferative disease-associated alteration in a Sal2 nucleic acid.

63. The method of claim 62, wherein the ability of said test compound to alter proliferation is measured by measuring the ability of a virus to propagate in said first cell contacted with said test compound, relative to said second cell not contacted with said test compound.

64. The method of claim 63, wherein said virus is a T-HR mutant virus.

65. The method of claim 62, wherein said first and second cells are mammalian cells.

66. The method of claim 62, wherein said first and second cells are in the same mammal or in different mammals.

67. The method of claim 66, wherein said mammal is a transgenic mouse.

68. The method of claim 66, wherein said mammal is a knockout mouse comprising a knockout mutation in a genomic *mSal2* gene.

69. The method of claim 62, wherein said first and second cells are ovarian cells.

70. A method of identifying a compound which alters cell proliferation, the method comprising:

a) exposing a cell or a cell extract to a test compound, and

b) measuring whether said test compound alters Sal2 levels, relative to Sal2 levels in a cell or cell extract not exposed to said test compound.

71. The method of claim 70, wherein said Sal2 is Sal2 protein.

72. The method of claim 70, wherein said Sal2 is *Sal2* nucleic acid.

73. The method of claim 70, wherein said measuring is by measuring Sal2 protein levels.

74. The method of claim 70, wherein said measuring is by measuring *Sal2* nucleic acid levels.

75. The method of claim 70, wherein said cell has a proliferative disease-associated alteration in a *Sal2* nucleic acid or said extract is from a cell having a proliferative disease-associated alteration in a *Sal2* nucleic acid.

76. The method of claim 70, wherein said exposing is with a cell and said cell is an ovarian cell.

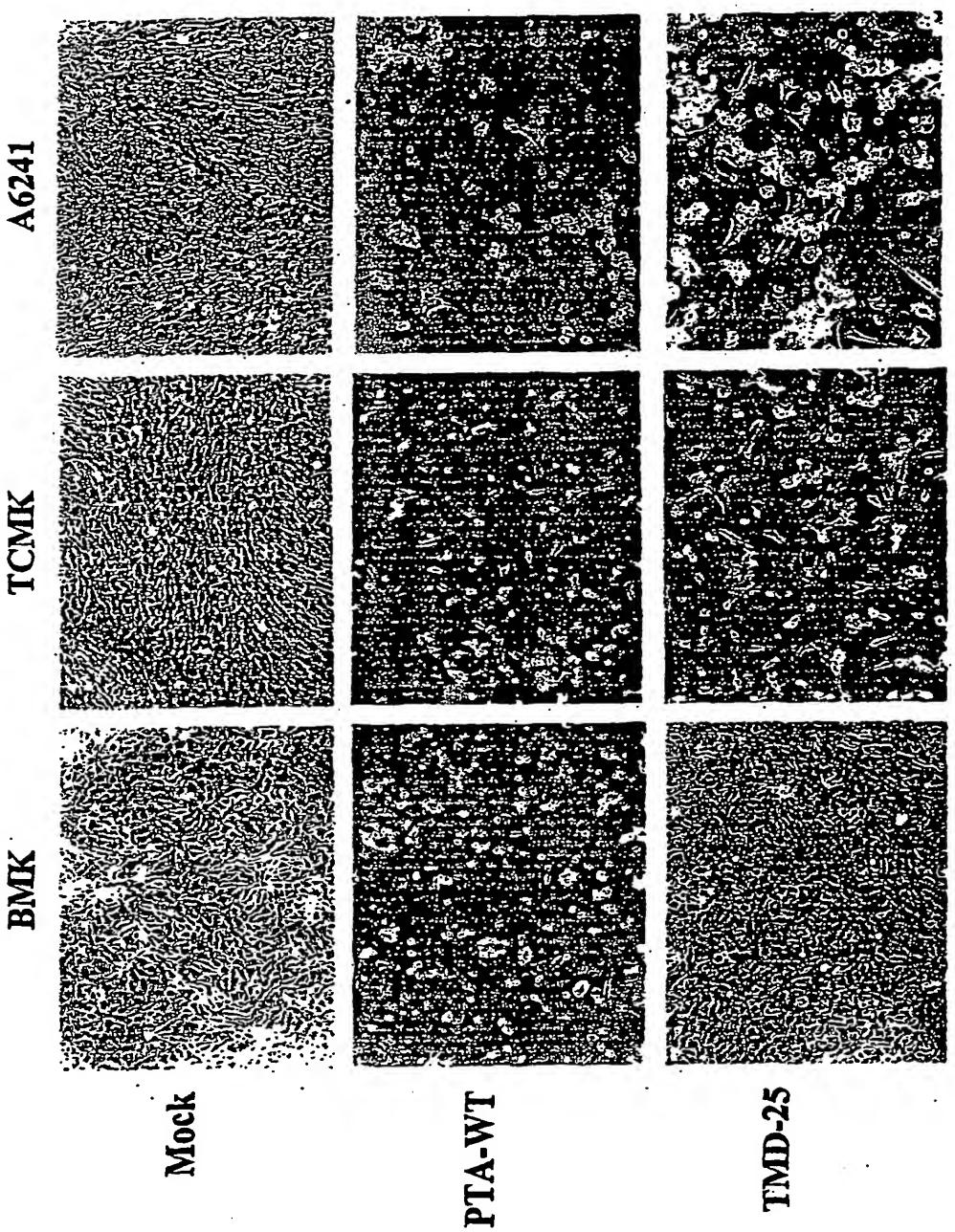
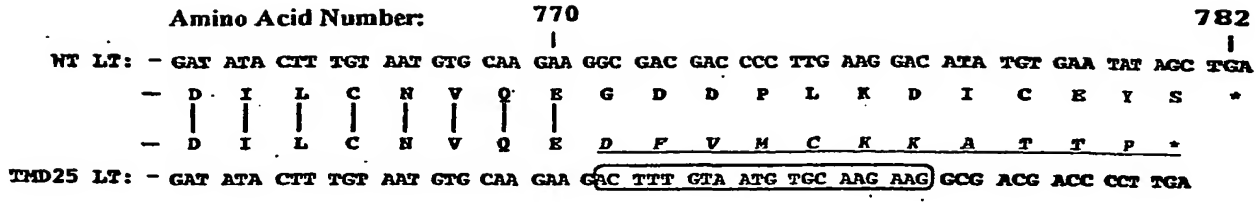
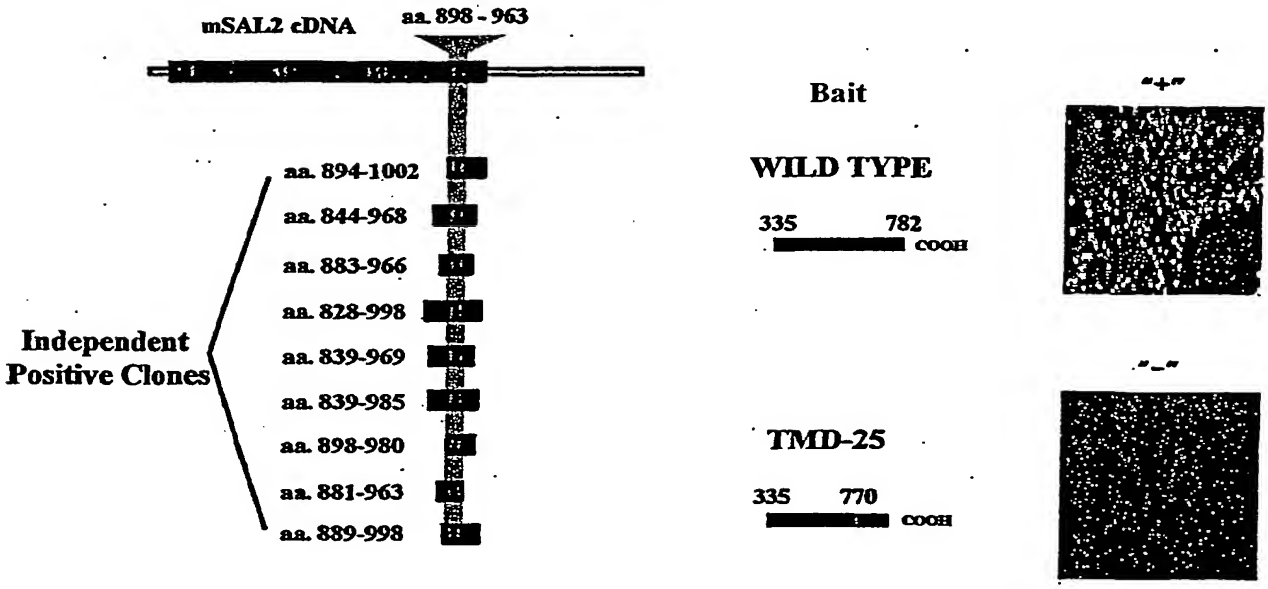


Fig. 1

A



B.



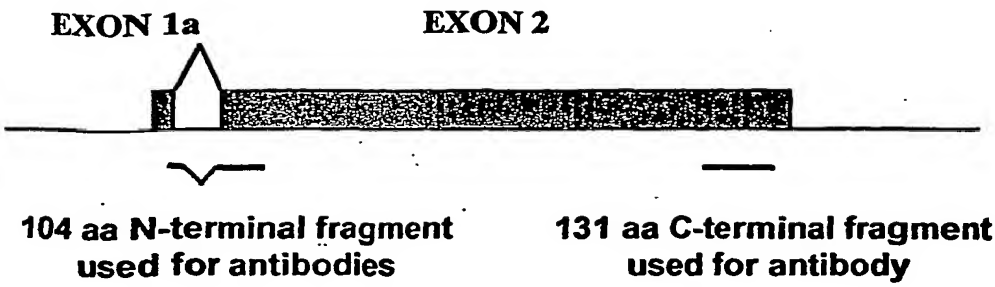
C.

Large T Deletions — Growth on His⁻ Plate

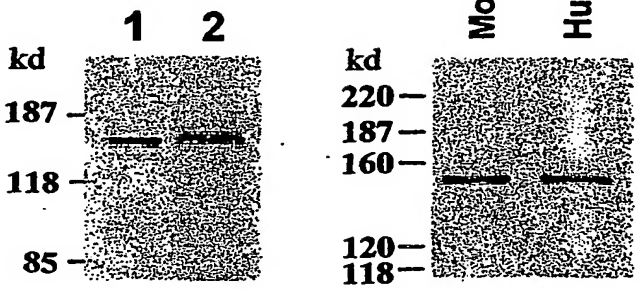
Wild Type	—	N	V	Q	E	G	D	D	P	L	K	D	I	C	E	Y	S	*	+
335-780	—	N	V	Q	E	G	D	D	P	L	K	D	I	C	E	*			+
335-776	—	N	V	Q	E	G	D	D	P	L	K	*							+
335-774	—	N	V	Q	E	G	D	D	*										–
335-770	—	N	V	Q	E	*													–
Δ 774	—	N	V	Q	E	G	D	D	_	L	K	D	I	C	E	Y	S	*	+
Δ 775	—	N	V	Q	E	G	D	D	P	_	K	D	I	C	E	Y	S	*	+
Δ 776	—	N	V	Q	E	G	D	D	P	L	_	D	I	C	E	Y	S	*	+
Δ 774-776	—	N	V	Q	E	G	D	D	_	_	_	D	I	C	E	Y	S	*	–

Fig. 2

A.



B.



C.

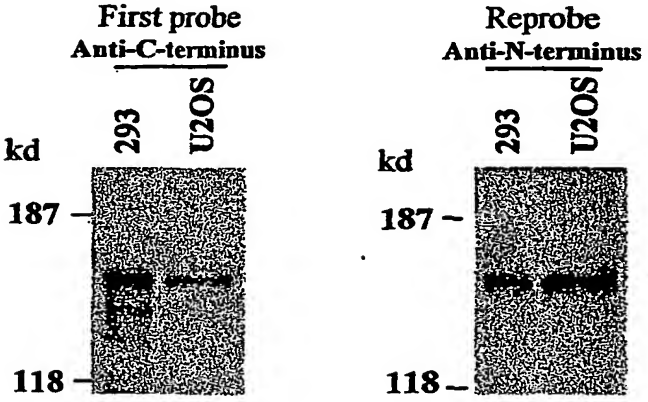
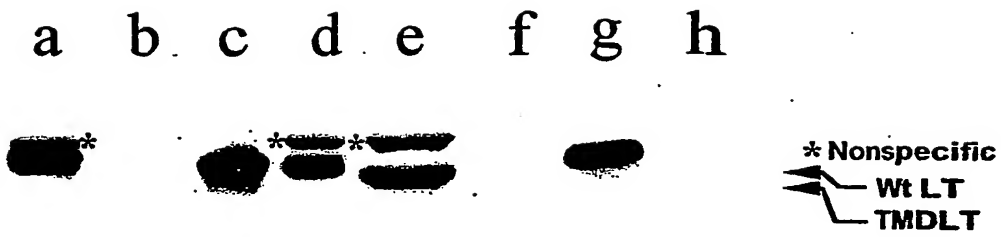


Fig. 3

A.



B.

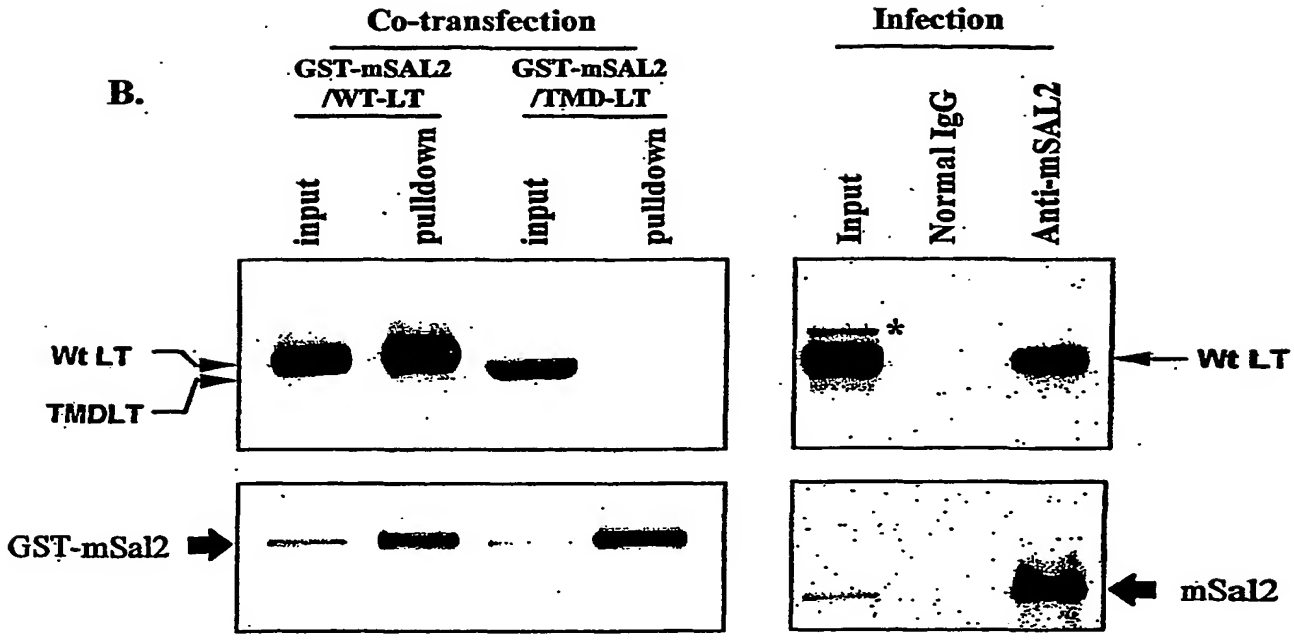
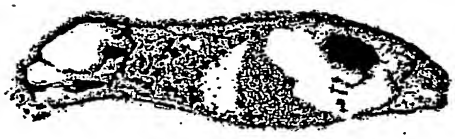


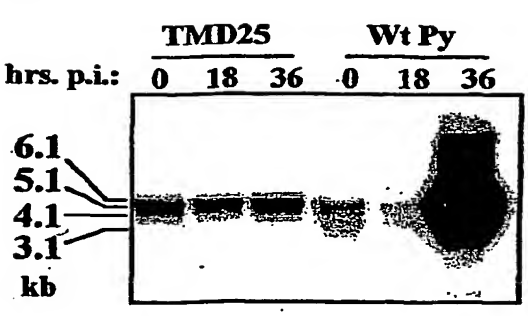
Fig. 4

A



TMD-25 WILD TYPE

B



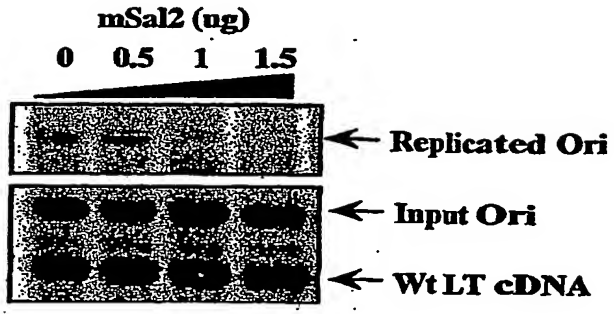


Fig. 5

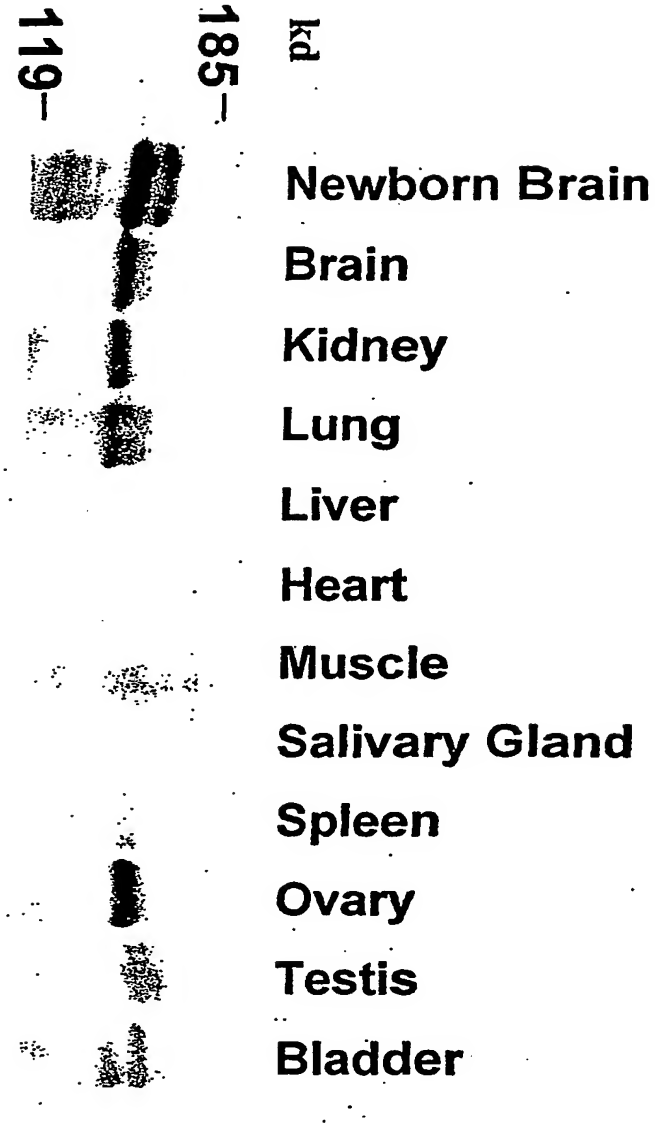


Fig. 6

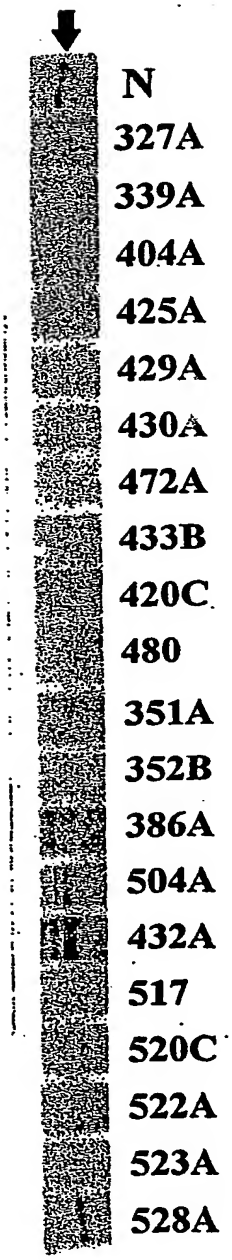


Fig. 7

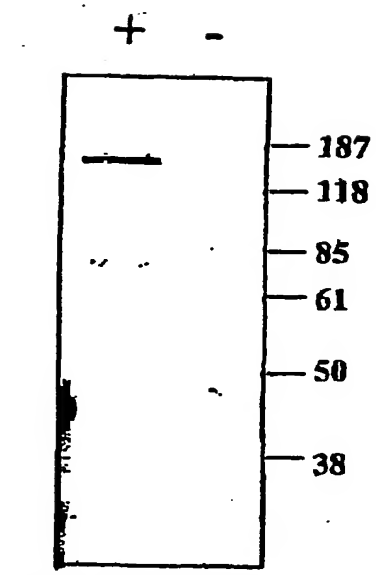


Fig. 8

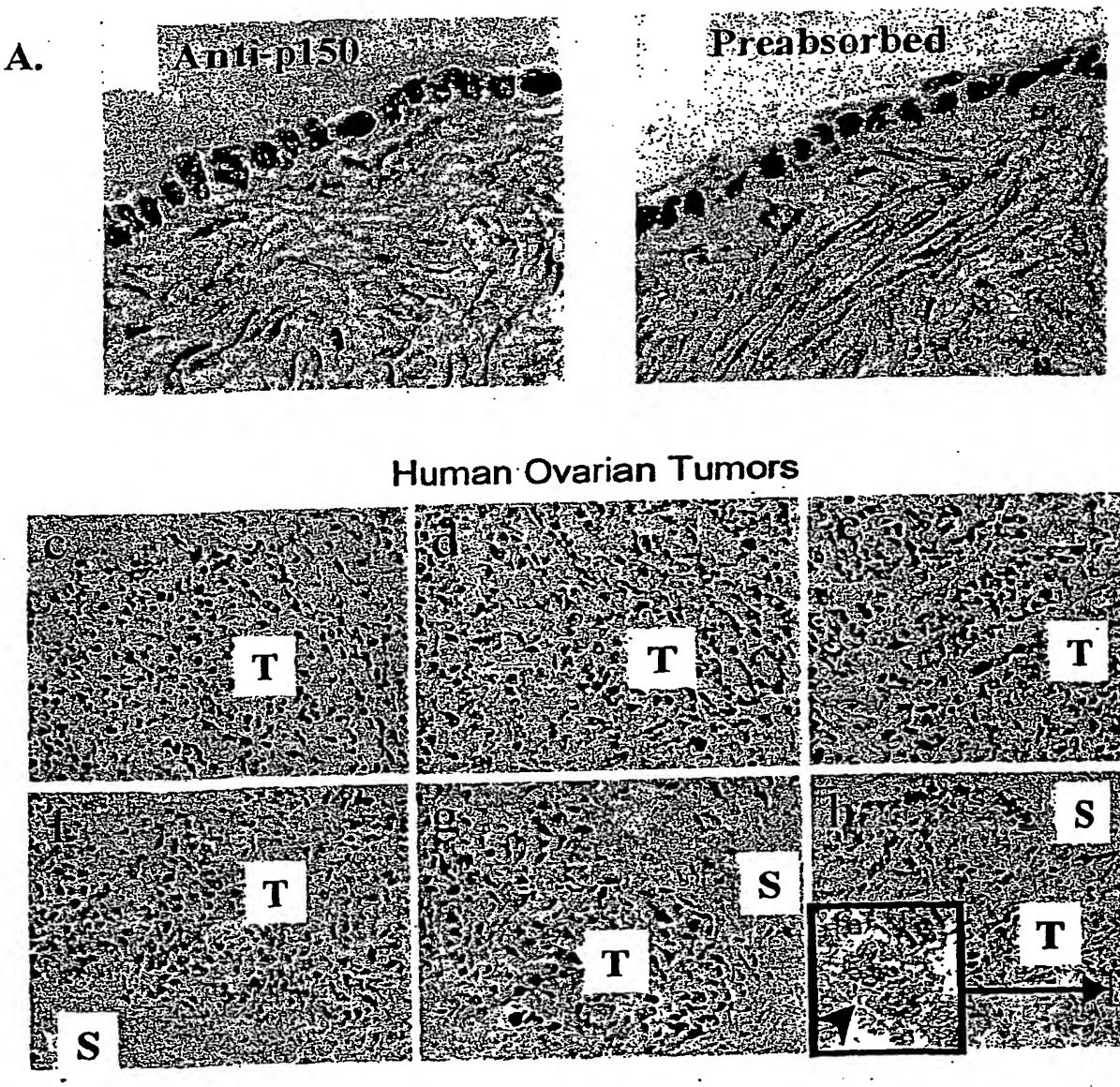
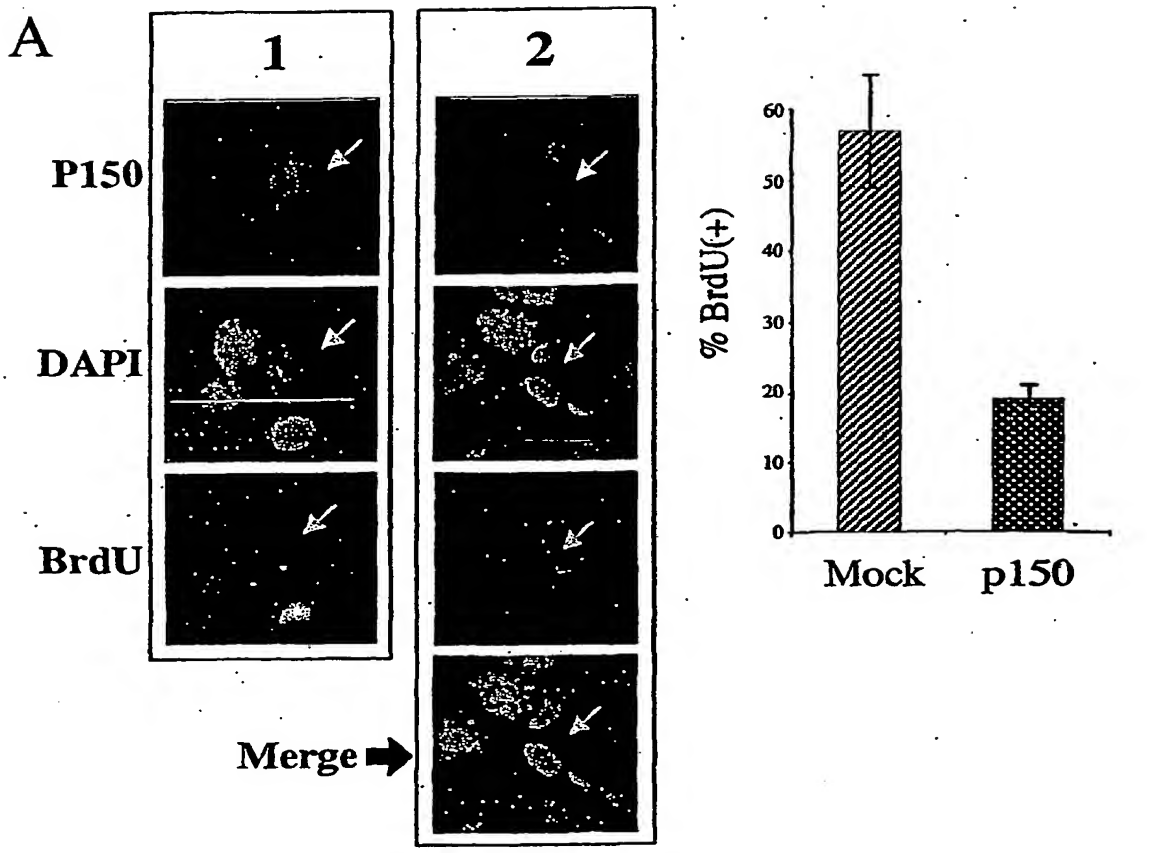

A.
Anti-p150
Preabsorbed
Human Ovarian Tumors
B.
c
T
d
T
e
T
f
T
S
g
S
T
h
S
T

Fig. 9



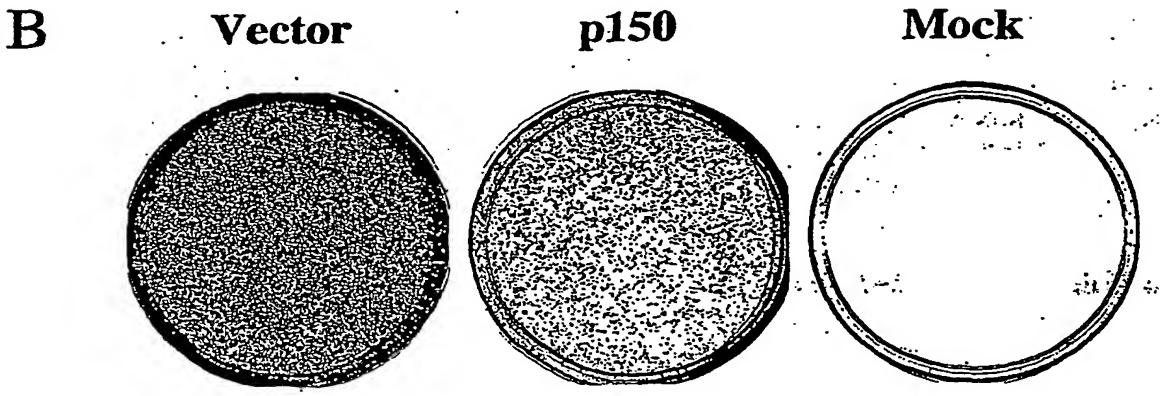


Fig. 10

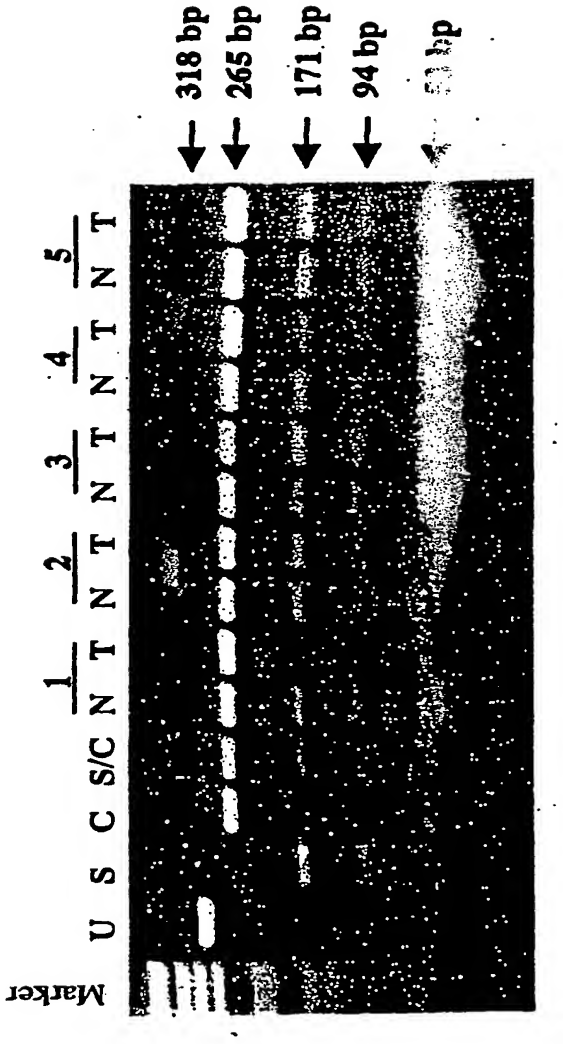


Fig. 11

SEQUENCE LISTING

<110> President and Fellows of Harvard College
The Brigham and Women's Hospital, Inc.

<120> Diagnosing and Treating Cancer Cells
Using Mutant Viruses

<130> 00742/062WO2

<150> 60/216,723
<151> 2000-07-07

<150> 09/812,633
<151> 2001-03-19

<150> 09/812,471
<151> 2001-03-19

<160> 21

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1005
<212> PRT
<213> Homo Sapiens

<400> 1

```
Met Ala His Glu Ser Glu Arg Ser Ser Arg Leu Gly Val Pro Ala Gly
 1               5                  10                  15
Glu Pro Ala Glu Leu Gly Gly Asp Ala Ser Glu Glu Asp His Pro Gln
            20                  25                  30
Val Cys Ala Lys Cys Cys Ala Gln Phe Thr Asp Pro Thr Glu Phe Leu
        35                  40                  45
Ala His Gln Asn Ala Cys Ser Thr Asp Pro Pro Val Met Val Ile Ile
    50                  55                  60
Gly Gly Gln Glu Asn Pro Asn Asn Ser Ser Ala Ser Ser Glu Pro Arg
65                  70                  75                  80
Pro Glu Gly His Asn Asn Pro Gln Val Met Asp Thr Glu His Ser Asn
                85                  90                  95
Pro Pro Asp Ser Gly Ser Ser Val Pro Thr Asp Pro Thr Trp Gly Pro
            100                 105                 110
Glu Arg Arg Gly Glu Glu Ser Ser Gly His Phe Leu Val Ala Ala Thr
        115                 120                 125
Gly Thr Ala Ala Gly Gly Gly Gly Gly Leu Ile Leu Ala Ser Pro Lys
    130                 135                 140
Leu Gly Ala Thr Pro Leu Pro Pro Glu Ser Thr Pro Ala Pro Pro Pro
145                 150                 155                 160
Pro Pro Pro Pro Pro Pro Pro Pro Gly Val Gly Ser Gly His Leu Asn
                165                 170                 175
Ile Pro Leu Ile Leu Glu Glu Leu Arg Val Leu Gln Gln Arg Gln Ile
            180                 185                 190
```

```
His Gln Met Gln Met Thr Glu Gln Ile Cys Arg Gln Val Leu Leu Leu
        195                 200                 205
Gly Ser Leu Gly Gln Thr Val Gly Ala Pro Ala Ser Pro Ser Glu Leu
    210                 215                 220
Pro Gly Thr Gly Thr Ala Ser Ser Thr Lys Pro Leu Leu Pro Leu Phe
225                 230                 235                 240
Ser Pro Ile Lys Pro Val Gln Thr Ser Lys Thr Leu Ala Ser Ser Ser
                245                 250                 255
Ser Ser Ser Ser Ser Ser Ser Gly Ala Glu Thr Pro Lys Gln Ala Phe
            260                 265                 270
Phe His Leu Tyr His Pro Leu Gly Ser Gln His Pro Phe Ser Ala Gly
        275                 280                 285
Gly Val Gly Arg Ser His Lys Pro Thr Pro Ala Pro Ser Pro Ala Leu
    290                 295                 300
Pro Gly Ser Thr Asp Gln Leu Ile Ala Ser Pro His Leu Ala Phe Pro
305                 310                 315                 320
Ser Thr Thr Gly Leu Leu Ala Ala Gln Cys Leu Gly Ala Ala Arg Gly
                325                 330                 335
Leu Glu Ala Thr Ala Ser Pro Gly Leu Leu Lys Pro Lys Asn Gly Ser
            340                 345                 350
Gly Glu Leu Ser Tyr Gly Glu Val Met Gly Pro Leu Glu Lys Pro Gly
        355                 360                 365
Gly Arg His Lys Cys Arg Phe Cys Ala Lys Val Phe Gly Ser Asp Ser
    370                 375                 380
Ala Leu Gln Ile His Leu Arg Ser His Thr Gly Glu Arg Pro Tyr Lys
385                 390                 395                 400
Cys Asn Val Cys Gly Asn Arg Phe Thr Thr Arg Gly Asn Leu Lys Val
                405                 410                 415
His Phe His Arg His Arg Glu Lys Tyr Pro His Val Gln Met Asn Pro
            420                 425                 430
His Pro Val Pro Glu His Leu Asp Tyr Val Ile Thr Ser Ser Gly Leu
        435                 440                 445
Pro Tyr Gly Met Ser Val Pro Pro Glu Lys Ala Glu Glu Glu Ala Ala
    450                 455                 460
Thr Pro Gly Gly Gly Val Glu Arg Lys Pro Leu Val Ala Ser Thr Thr
465                 470                 475                 480
Ala Leu Ser Ala Thr Glu Ser Leu Thr Leu Leu Ser Thr Ser Ala Gly
                485                 490                 495
Thr Ala Thr Ala Pro Gly Leu Pro Ala Phe Asn Lys Phe Val Leu Met
            500                 505                 510
Lys Ala Val Glu Pro Lys Asn Lys Ala Asp Glu Asn Thr Pro Pro Gly
        515                 520                 525
Ser Glu Gly Ser Ala Ile Ser Gly Val Ala Glu Ser Ser Thr Ala Thr
    530                 535                 540
Leu Met Gln Leu Ser Lys Leu Met Thr Ser Leu Pro Ser Trp Ala Leu
545                 550                 555                 560
Leu Thr Asn His Phe Lys Ser Thr Gly Ser Phe Pro Leu Pro Leu Cys
                565                 570                 575
Ala Arg Ala Leu Gly Ala Ser Pro Ser Glu Thr Ser Lys Leu Gln Gln
            580                 585                 590
Leu Val Glu Lys Ile Asp Arg Gln Gly Ala Val Ala Val Thr Ser Ala
        595                 600                 605
Ala Ser Gly Ala Pro Thr Thr Ser Ala Pro Ala Pro Ser Ser Ser Ala
    610                 615                 620
Ser Ser Gly Pro Asn Gln Cys Val Ile Cys Leu Arg Val Leu Ser Cys
```

625 630 635 640
Pro Arg Ala Leu Arg Leu His Tyr Gly Gln His Gly Gly Glu Arg Pro
645 650 655
Phe Lys Cys Lys Val Cys Gly Arg Ala Phe Ser Thr Arg Gly Asn Leu
660 665 670
Arg Ala His Phe Val Gly His Lys Ala Ser Pro Ala Ala Arg Ala Gln
675 680 685
Asn Ser Cys Pro Ile Cys Gln Lys Lys Phe Thr Asn Ala Val Thr Leu
690 695 700
Gln Gln His Val Arg Met His Leu Gly Gly Gln Ile Pro Asn Gly Gly
705 710 715 720
Thr Ala Leu Pro Glu Gly Gly Gly Ala Ala Gln Glu Asn Gly Ser Glu
725 730 735
Gln Ser Thr Val Ser Gly Ala Gly Ser Phe Pro Gln Gln Gln Ser Gln
740 745 750
Gln Pro Ser Pro Glu Glu Glu Leu Ser Glu Glu Glu Glu Glu Glu Asp
755 760 765
Glu Glu Glu Glu Glu Asp Val Thr Asp Glu Asp Ser Leu Ala Gly Arg
770 775 780
Gly Ser Glu Ser Gly Gly Glu Lys Ala Ile Ser Val Arg Gly Asp Ser
785 790 795 800
Glu Glu Ala Ser Gly Ala Glu Glu Glu Val Gly Thr Val Ala Ala Ala
805 810 815
Ala Thr Ala Gly Lys Glu Met Asp Ser Asn Glu Lys Thr Thr Gln Gln
820 825 830
Ser Ser Leu Pro Pro Pro Pro Pro Pro Asp Ser Leu Asp Gln Pro Gln
835 840 845
Pro Met Glu Gln Gly Ser Ser Gly Val Leu Gly Gly Lys Glu Glu Gly
850 855 860
Gly Lys Pro Glu Arg Ser Ser Ser Pro Ala Ser Ala Leu Thr Pro Glu
865 870 875 880
Gly Glu Ala Thr Ser Val Thr Leu Val Glu Glu Leu Ser Leu Gln Glu
885 890 895
Ala Met Arg Lys Glu Pro Gly Glu Ser Ser Ser Arg Lys Ala Cys Glu
900 905 910
Val Cys Gly Gln Ala Phe Pro Ser Gln Ala Ala Leu Glu Glu His Gln
915 920 925
Lys Thr His Pro Lys Glu Gly Pro Leu Phe Thr Cys Val Phe Cys Arg
930 935 940
Gln Gly Phe Leu Glu Arg Ala Thr Leu Lys Lys His Met Leu Leu Ala
945 950 955 960
His His Gln Val Gln Pro Phe Ala Pro His Gly Pro Gln Asn Ile Ala
965 970 975
Ala Leu Ser Leu Val Pro Gly Cys Ser Pro Ser Ile Thr Ser Thr Gly
980 985 990
Leu Ser Pro Phe Pro Arg Lys Asp Asp Pro Thr Ile Pro
995 1000 1005

<210> 2
<211> 16080
<212> DNA
<213> Homo sapiens

<400> 2

atatcacacc ccagctggct atgtaatcat gaaataagga gaaacacata aatatttggt 60
taaaacacct ttaatgatag agggaaagac actaatatct cccgtctgtt cttgacattt 120
tactaggtta ggaagctctg gagcctacag cttgaggaga agccatcgtt caagtcagtc 180
aatagcaaaa ccctcactct ctcctcctca gaactcctgt tccaaatgat cctatgttaa 240
gagtaaatac tacaactcat tacaagacgg agaggcaggg aggacgccac ctggagctgg 300
gactcttaag aaccagacaa tgacaaagac acaagcccca gcctacggat aggcaaaatg 360
ggtaggggtc ttgaaagagg aagataagga aaatacaagg ggccagggaa taaaggaggg 420
agttatctaa aactagaagc atactagtgc taggaaatcc cccatgatcc ctggtacacc 480
tctgcacact atgtcactat tagcccaaaa gaatattaac gagaatgtcc acattcacaa 540
gaatttgagg ccttttccct tacatcatgt ccctttctta gtcacatagg taccagcaag 600
ccctatgttc tagcaacatt ccttaactct ctcatcatta gttcatcaac catgctgacc 660
aaaaatgctc cttaaagata cgaacttcac atttcccaaa tatctcctgg gagacctctt 720
ggcaagaaat cagcttgttt cccaactttg agaggtcatc atgaatgaga agctggagag 780
gtcttggcac actgaccagc caaaaccttt accttaatgt gaccatcagg ggatttactg 840
ggaaaatttt cctatgccct tccttcattt ctccctactt cctagggttg ggtcaccaat 900
tactggagca tcttcagtac cggcaccttc tggagcaggg ggaggaagaa ggaatgtaca 960
gtttgctact tcttgtctat gatgggcttc tcaggcactg ccttgggtgc aggaggctga 1020
aataggaggg gggctgtctt ctccttggct tccctggatc ccattgttgg aggcaccttc 1080
ccagccacag ttcctaggcc aaacagcact ggtggggcca ggcttggagt ggtagtggag 1140
gtggagctgg aattccaggg cttcatgggc aggccatttg acaggaatgc cacatactgg 1200
ttctagaaag ataggggacc catacccacc agctgagcag aaaggtcacc ccagaggagt 1260
ggcactgggc cctccagaga cagctgccag ccctttttgg ctaggctgca atgccaaatg 1320
taggtgctca ggtgcaccta ccaaagggaa agggagagga gagaggaggg ggaagaaggg 1380
tcacaccagg gaagctggag agggttcccc ttgagaaagc tgcagagaat ctatgttcct 1440
caggtacaaa gaatgaggag ggaagaaaaa ttccttaggg ggccatcccc ttgtaagcac 1500
agtaatttcc aagctcaggg actacagaaa agccactagg gacataacat gttaagaact 1560
tagagaaaaa gacaaaatca gggctcataa ctctgggagg tccttttgtg aagctgtttc 1620
tgctctgtgg gacaaagagc agcaggtaca gaaaaacagg ctcatgggat cgtggggtca 1680
tcttttcggg gaaaggggga gagccctgtg gaggtgatgg aaggcgaaca gccagggact 1740
agagaaagag cagcaatatt ctgagggcca tgggggggcaa agggctgtac ctggtggtgt 1800
gccaggagca tatgcttctt gagggtagcc cgctcaagaa agccctgcct gcagaaaaca 1860
caagtgaaga gcggcccctc cttggggtgg gtcttctgat gctcctccag agctgcctgg 1920
gagggaaagg cctggccaca cacttcgcag gcctttctgc tgctgctctc tcctggctcc 1980
tttctcattg cctcctgcag gctcagctcc tctaccaagg tcacgctggt ggcttccctt 2040
tctggggtga gtgctgatgc cggacttgag cttctctccg gtttgccccc ctcttccttg 2100
cctcctaaaa caccactgct tccctgctcc attggctgag gctgatccag gctgtcaggt 2160
ggtggtggtg gtggcaaaga agactgttga gtagttttct cattactgtc catctccttc 2220
ccagctgtgg ctgctgccgc cactgtcccc acctcctcct ctgccccaga tgcctcttct 2280
gaatcacctc tcactgatat tgccttctca cctccactct ctgagcctct ccctgccagg 2340
gaatcttcat cagtcacatc ttcctcttct tcctcatcct cctcttcctc ctcctcagac 2400
aactcctctt ccggtgatgg ctgctgggac tgctgctggg ggaaactccc tgccccggag 2460
actgtagatt gctcggagcc attctcctga gcagctcctc caccttcagg gagtgcagta 2520
ccaccgttgg ggatctggcc ccccaggtgc atccggacat gctgctgcag agtgacagca 2580
ttggtgaact tcttctggca gatggggcag gaattctgtg cccgggcagc tggactggcc 2640
ttgtggccca cgaaatgtgc acgcagatta cccctggtgg agaaggctct gccacacact 2700
ttgcatttga agggcctctc acctccatgt tggccataat gaaggcgtag ggcccgagga 2760
cagctaagca ctcggagaca gatgacacac tggttaggtc cagaagaggc tgaggatgaa 2820
ggtgcagggg cagaggtggt gggggctcct gaggcagctg aggtcaccgc cacagctcct 2880
tgccggtcaa tcttttctac cagttgctgc agctttgatg tctcagaggg tgaggccccc 2940
aagggctcta gcacataggg gaaggggaag ctgccagtgg acttgaagtg gttggtaagc 3000
agtgcccagc ttggtagtga agtcaccaac ttacttagtt gcatgcgagt tgccgtgcta 3060
ctttctgcca ctccactgat ggctgagccc tcactccctg gggggtgttt ttcatcagct 3120
ttattcttgg gttccactgc tttcatgagc acaaacttat tgaaagcagg gagtcctgga 3180
gccgtggctg tgcctgcact ggtggagagc agagtcaggc tctctgtggc actgagtgct 3240
gttgtggagg ccaccagagg cttgcgctca acccctccac ctggagtggc tgcctcctcc 3300

tcggccttct ctggtggcac ggacatacca taaggcaagc cactgctggt aatgacatag 3360
tctaggtgct ctggtactgg gtgtgggttc atctgcacat gtgggtactt ctcacgatgc 3420
cggtggaaat gcactttgag gttgccacgg gtggtaaaac ggtttccaca gacattgcac 3480
ttatagggcc tctcacccgt gtgggaacga aggtggatct gcagggcact gtcactgcca 3540
aatactttgg cacagaagcg gcatttgtgc cttccaccag gcttctccaa gggacccatc 3600
acttctccgt agctcagctc accacttcca ttctttggct tcaggagccc tggggaggca 3660
gtggcctcaa ggcctcgggc tgccccaaga cactgtgctg ccagtagtcc cgtggtgctt 3720
gggaatgcca gatgaggcga ggcaatcagc tgatctgtgc tgcctggcaa ggctggggaa 3780
ggggcagggg tgggtttgtg gcttcgccca acccctccag cagagaaagg atgctgtgac 3840
cccagtgggt ggtaaaggtg gaagaaggcc tgcttgggcg tttctgcccc tgaagaggaa 3900
gaggaggagg aggaggaaga tgccagtgtc ttgctggttt ggacaggctt gatggggctg 3960
aagaggggta gtaggggctt ggtggaagag gcagtccctg tcccaggtag ctctgaggga 4020
ctggcagggg cacccaccgt ctggcctaag gagccaagca acagcacctg cctgcagatt 4080
tgctcagtca tctgcatctg atggatctgc cgctgctgca gcacccgtag ctcttccaag 4140
atcaggggga tattcaagtg gccactgcct acccctgggg gcggaggggg tggtggagga 4200
ggaggggtg caggggtcga ttctggaggt aatggggttg ctcccagctt gggactggcc 4260
aagatcaggc ccccgcctcc cccagccgct gtacctgtgg cagcgaccag gaaatgccct 4320
ggagactcct ctcctctcct ctctgggccc caggtgggat ccgtgggcac ggaggaccca 4380
gaatctgggg ggttgctatg ctctgtgtcc atgacctgag gattattgtg accctcaggc 4440
cggggttcag aggaggccga agagttgttg gggttctcct ggcccccaat tatcaccatt 4500
acaggagggt cagtagaaca tgcgttctgg tgggcgagga attcagttgg gtcagtgaat 4560
tgtgcgcagc acttggcaca gacttggggg tgatcctcct cgctagcatc acctggggag 4620
aagacaagga gagagagcgt gggtggcgca gttgggttgg gtataccgag gctctaatta 4680
acaaggaggc cagtaaccgc tagttggggg tggggagatg agctcaccat cagggccatg 4740
cagaagtcta gagctcaggc ctgatccgtg tggacaggag acaacccggc atggggcagg 4800
ggggtgggga gggaggaggg gaggggggca agagcatgct actcccctcc tcagccaccc 4860
tcccttcccc aggccacaag cgagttcacg gaataggtgt ggggacaggg gcctacgcag 4920
agaatcatgc attttctccc acccaccgaa agtcttcgcc gcccctgcgc atccccctcc 4980
gcccccaccc ctgcccagcc cgaccgaccc taccgcacct ccgagctctg ccggctcccc 5040
gcagggcacc ccgagacgag agctcctctc ggattcgtgc gccatggttg tgggggaagt 5100
ggagggccag gtggggtggg agacaatgga tattgggatt gagggaggcg atggccgctg 5160
ggtctgcggc agcctctgca cccagcggcc cagactgcgg agatggagat cggcagcggc 5220
gggggcaggg agcagcggcg gaggggggagg ggagcgagga ggcggggaga agctggagtg 5280
agaaagcggg gagaggggag atctgggagg agctgatgag gaggggagtt tatggggagg 5340
agctgctggg gagggaggcg ggagctagag gaggcgggag aaggggagcgc tagcgggggc 5400
gtgggggcgg gagctcagag ctcgggagag tttccggagg cgcagtgaca ggtgctgtga 5460
agcactgcgg gggtccacct ttcccggtcc ctggccagct ccccccatct gcagatgcct 5520
ttgcccaggc ctaccctcct ccccccgccc tcccctccta agctctaggg gcacagtggg 5580
aaacgtagcc ctgctcagtg gagcaaggcg ataggcttct cttatttttc tttggataaa 5640
ggatccgctg agcttggaaa aagtggattc cagagagggt cgtctgatct cctcagaggt 5700
ctgagggcca gaagaagagg gggagatcag aacatccact cctcaccagc acacacaccc 5760
caaaatattc gaagttttgt ctcgtctttc tcacttccat tcccacccta cccccatccc 5820
tctccacaaa agaagtttct cagggtgggc ggctgcaagg tagaatttcc caggaagtca 5880
tttcaggact ctctgcggaa cactaagccc cttcactccc cgcccctcct cccctgaat 5940
aatagctgaa tgcaggttac tccgcagatc gcccagccta cacaacacct aattcataga 6000
gtccatgctt atttaataag ccatctccta tttagtaccc tcttcctcct ctattctcct 6060
cttgcaacat tcctcacacc gtcactatta aagacagtgg gtttggggag acgctagcct 6120
gcagaggcct acggaggccc acccagctct aacctggggg ggaggggagc cctcttgaaa 6180
caatgcggta ggaactacca ggcagccctc agtgtctaaa gccctttcag ccccagcctg 6240
atttgaatgc ttagaaatag ctaacacctg ctcaccatca cagaggcagc ctcctattca 6300
gacaggataa gtaagaataa aatgcctcct ggaccaggta ttctggcatt ctctttttta 6360
ccttgaaatg agtcttaaag tgcttcccac ttcctaaaat actttctctt acatgcagga 6420
agtgaccaca agtccttggt tttgtggttt ccctgggcat cagtaaacct aaattgtttt 6480
aatcccagtt ctattcttgc ctcactgata aaactgagac atggtggtca gtcacaccat 6540
gttataccac cgtttccctc ttcataaagt ggtaatattg tagctgcagt attttactca 6600

gaaaaatatt gtggggacaa aaaattgaaa aattggacaa tattaatgtg taaaccaggt 6660
atggtggtgc acacttgtag ccccagctac ttgggaggct gaggcaggag aattgcttta 6720
gtccaggagt ttgaggctgc agtgagctgt gatcacacct gtgaataacc actgccctcc 6780
agcttcggca acatagtgag gccccattac tttaaaaaaa aaaaaaagcc gggcgcggtg 6840
gctcactgta atcccagcac tttgggaggt gggcagatca cgaggtcaga agttccagac 6900
cagcatggcc aacatgttga aaccccgtct ctactaaaaa tacaaaaatt agctgggcat 6960
gatggtacac ctttaatccc agctactggg gcagctgagc caggagaatg gcttgaaccc 7020
aggaggtgga ggttgcaggg ggctgagatc gtggcattgc actccagcct gggcaacaag 7080
agtgaaactg cgtctcaaaa aaaaaaaaaa gtctaaaaaa attaatatgt acatgtgaga 7140
tttttaaagt ttggggagtc ctgaatttaa tcaatgagat aatttacatt gtcagtagca 7200
aaataatcga agtaacctta aatacacata tactaaaatt agatctgttt tccatgttgt 7260
ttgttaatct tattaatttc tgaggtaaga tattggctaa tatcagcagc atatttcaaa 7320
ggtaggaagt cttttattgc agtgggtggg ggagctgaaa caacctattt aaaatattag 7380
taacatccac tttacttctc aacataaatt ttgcctgtgt ttttaaactt aaaacagttt 7440
actgaattat gttttgaaac ttcagataat aaggctctta gcattgtgag tcataattct 7500
gaaatggacg ggttctgtgc ttccaggcct ggacttacaa atgagggagg ggggttctat 7560
ttcagtttat ggcaagtcac agttttgtgc aatgtggttt atttttacag ataaggaaac 7620
tgaagcttgg agaggttaag tgacttttcc aagttcacac agtaattcag tgaagcaagc 7680
attcagaatt ttgactcctg tccaatgctt tctcaagcac atcaactttg tatggcttcc 7740
ctaatgctag agaaagggcc ctgtgtggct tctacctgcc atttgctccc tggccttagt 7800
cagggagagg gaatcagatg gaggctttct actgagcatt tgttaattag cattgaacat 7860
ttgatatcca gttgctgttt tgtcaagtct tctgacaaga aaagaaatcc ttttcttttc 7920
atcttctcct gggaaacact gtccccttcc ttgctcttta atgaaatgtg ctttctgatg 7980
cgtaatttga tctaagctct tctttaaggt aaatttagtc cctggtgaaa ggtgactgga 8040
tcaacagcca cctgtaagag gaaccctcca tttctcagta ctttgcactc actgcacatc 8100
ctgaaaaggg gggcaggatt cttacacaaa catgaatgaa gtcacaaatg caggaataaa 8160
ctaaactggt aatggtgtcc ctagatagca gataaggtga ggtaagctat ctccggtcaa 8220
atgcaaagtc cggggtggga ctaagacctg gacaagcttg tttaaactta tagagagctg 8280
aaatgacaaa gaaaagggaa accaggtggc ttcccttcta aatctagtgt cccatcagat 8340
tgcttcttta ggcttcagag agaactgttc gggagaacaa agagaaaaat aggtgagttg 8400
tatgtagcag ggtgatacat ttgaacagcg gttttcaaat tttgctgccc attaggatta 8460
ccagaagaga gttttaaaat ttttatgttt aggtgcagtg gtctgttcct tgtagtccca 8520
gctactctgg aggctgaggc gggaggatca cttgagctca agggtttgag actccatctc 8580
aaaatctcaa aaaaaagaag aaaaaaaaaa gaaaagaaag tttaagcaca gtgggtacct 8640
catgcctata atcctagcac ttttggaggc caaggcagga ggattgcttg aggccaggaa 8700
tttgagacca gcctgcataa catagtgaga cccccatctc tgcaaaagca accaaccaac 8760
caaacttaaa aaaaatccct gtgtccaggc cacatcccag gctaattaat tcataatccc 8820
tgaggatagg atccaggcat tagtttgata aagctcctca ggtgattcca atgagaatac 8880
aaagatggtg acacaatgat gagacccaca tggaggactg ccctttccat catacccttcc 8940
accctgctcc tcacagatct tacctgagct aaacttggcc acaattggga cacagacaaa 9000
atgaactctc aatgctaaat cttcccatca ggtcccctcc ctacagtgcc cacaaccaca 9060
cattaacttc cttgtatcct ttcccagtga aaaatctgct tccatgaata gaatttgata 9120
taatttacac cttactgtaa gtttaagtga ttgcatttct ttcccaggta tgggtatctt 9180
gaagcatatt tttttctttt ttaattgata tttgagccat atttcttttt ttttctttct 9240
ttttttttt ttttttttt ttgagacgga gttttgctct cattgcccag gctggagtgc 9300
aatggcatga tctcggctca ccgcaacctc cacctcccag gttcaagcga ttctcctgcc 9360
tcagccttcc caagtagctg ggattacagg catgtgccac caagcccggc taattttgta 9420
tttttagtag agatggggtt tctccatgtg ggtcaggctg gtctcaaact cccgacctca 9480
ggtgatctgc caacctcggc ctcccaaagt gttgggatta caggcgtgag ccatcgcgcc 9540
cggccaccat atttctaatt gtaaggtgaa aggctttgtt ctacagagtt caagcatcat 9600
ccacccatta aggctggagt gaagtggcac aatcatagct cactgcagac tctacctccc 9660
aggcttaggt gatcctccca cctcagcttc ctgaatagct gggactacag gcatgcacaa 9720
tcatgcccag ctaattaaaa tatttttttc tgtagagatg aggtttcact atgttgccca 9780
ggctgtctgg aatacctggg ctcaagggat cctcctgcct tgtccccaca aagtgctgag 9840
agtacagatg taagccactg cctctggccc acttacttat tattgacact gaacaatgct 9900

aattggtagc ttccataatt atgaattgat tctgtaacta ttgctactga ctacttctta 9960
gggaaatatc tcatcttctc ctccttactc ctctttccta aatgtagaca cataataatc 10020
ctttgcaacc cagacctact aatgtaacta tggcctatgt aacacagtag actaacaggc 10080
acaatgattg gtacacctgg tgctaagtga gaaaaagata tttgtttcca gaacaggaat 10140
atcttagatc aaacataaga atgttctttt aatgaaaatt tctttgactt caaaggactc 10200
aacacttaac atggaattca taccattttg gagctgggac ttcagagatc tgacactctc 10260
attgtcattg tgcacagtga ttcagacctg agttaaagtc ccagctctag aacattctaa 10320
tatttgtgat cttgggaaaa tttcttaatc tctcccagag tttgttttct tatttttttt 10380
tgggacagag tttcactctt gttgcccagg ctggagtgca atggcacgat cttagctcac 10440
cgcaacctcc gcctcccagg ttcaagcgat tctcctgcct caccctccct agtagctggg 10500
attacaggca tgtgccacca cgcccggcta attttgtatt tttttagtag agacggggtt 10560
tcttcatatt gctcaggctg gtctcaaact cccagcctca ggtgatctgc ccacctcggc 10620
ctcccaaagt gctggattac aggcatgagc caccgcgcct ggccagcctt ttttttttt 10680
ttgagacgga gtctcgctct gtcgcccagg ctggagtgca atggtcgccc aggctggagt 10740
gcaatggtgt gatctcggct cactgcaatc tccgcctcct gggttcaaac gattttcctg 10800
cctcagcctc ccaagtagct gggattacag gtgtgcgcca tcacacccag ctaatttttg 10860
tatttttagt agagatgagg tttcaccttg ttggccaggc tggtcttgaa ctcctgacct 10920
caagtgattt gcccacctca gcctcccaaa gtgctgagat tacaggcatg agctgctgtg 10980
cccggctgat ttctcttctt taaaatgagg gtactgccat acaaaggaag gaaattctga 11040
tacatgctac aacatgaatg aactttgtaa acattatgct ttcagacaaa tttgacttta 11100
attgagaaaa aaagagaaaa catactaagt gcaataaagc agacacaaaa ggacaaatat 11160
tgtatgattc cattagtatg aggtacccaa acattatatg agtccattaa tatgaaattt 11220
ggcaaggtca cacatacaga aagcagagta gaggctaaca gggctaaggg aatgggagaa 11280
tggggattta ttgtttaacg gttacagttt ctgtttgatg atgaaaaaga tattgaaaca 11340
gcagtaatgg ttacataaca tagtgaatgt acttaatgcc actgaattgt acacttaaaa 11400
atggttaaaa tggtaaattt tattacacat attttacaat aaaaaaattt tagccaggtg 11460
tggtggcatg cacctgtaat cccagctgtt caggaggctg aggcaggaga atctcttgaa 11520
ccctggaggt ggaggtttca gtgagccgag acgtgccact gcactccagc ctgggcaaca 11580
gagtaggact tggtctcaaa aaaagaaaaa aatttttttt gtaataataa gggagttggg 11640
gctgggcgtg gtggctcacg cctgtaatcc cagcactttg ggaggccaaa gtgggcggat 11700
catgaagtca ggagatcgag accatcctgg ctaacacagt gaaaccctgt ctctactgaa 11760
aatataaaaa attagccagg tgtggtggcg ggcgcctgta gtcccagcta cttgggaggc 11820
tgaggcagga gaatggtgtg aacccgggag gcggagctcg cagtgagcca agatcgcgcc 11880
actgcacccc agcctgggcg acagagcgag actccgtctc aaaataataa aaataaataa 11940
ataaataaat aaaataataa taataacgga gttgggagga aaaagaggaa atgcaaaaag 12000
ggcctagcac agtacctgaa tgctccacaa atattagcca tgggtgttag ttattatttg 12060
aatgtcaaaa gctgaatgaa gccctggggt aagaaaggtc acatgtgccc aaggtcacat 12120
agcttcaagg tccacactag attgaaaacc aagttttctg ttttcttatc tagtactctg 12180
taacaccagg actgagatac tctctattcc aaaatgtgtt ttttctgatc tgggaatacc 12240
taggttgagt ggcccaggga tcaataacct gagagatgag gctctttact tccaaatgta 12300
aacagagccc ccaaaactct acctttgcct tctttcctct cttgctgttc ttgctatctg 12360
ccaacttcca tctaaagtac tcccctctct ccctctagat ctgtttggct gctgtcctgg 12420
tttcttcttc tcactaaata tctgggtttc tgattgtttc ctttatttcc cagatgtact 12480
ggtttgcatt tttcccccag tcacatcctt tgtgttctct aatccagatt tctagactct 12540
gtaggggaga gagaaggttt tttttttcc tctctagagt ttttaagtga atagagtatt 12600
tcctgcccat cacttatatg caataactgt tctgttaggt tttgatgctc tggttaggga 12660
agctgagcaa aaacggctgg aaaacagatt tttcagactg tttcttggtg atgtcttagg 12720
tcactgcaga attttggctt ttaaaatatg taacaaaggc tcagcatttg catgttgtat 12780
atggcacata ttgcttacaa gaaggcaaaa gactcctgga aacattactg gcaccctaga 12840
ctactgacta aatgtcttct gatactcatg atgatatcca taatttcaca ggtacaccaa 12900
aggatacatg tgcccctaaa taagagccct tcctccctaa ctgtggagca tgctctgggg 12960
tagaaggaag tcagatgcct gaagatcaca taagtgaata gaaaccctgt ctataaaaaa 13020
ttagggaaaa ggagagctct cattctgttt tgcagaatgg atgctgcccc attcatgatt 13080
aagaaaattt attaatttaa aagaaaacca gaaaatgtga aatttatata ttataagctt 13140
ataagatcca ggaggaattt tagatacgat caaatagagc cacctcattt tgcagatgag 13200

gcccaatgac atccagatca taagtagcct aggatctttc actccagggg aattctgatg 13260
agaaaatcct taggctttct tacggtagat cttaacagag ggtgctactg cttccttgct 13320
ccttacattt gttcctgcct ttcatagctc aaaggcaaat tttcatcaaa aatttgttga 13380
tgccattggg tttaaacctt tactgtttct atggggatgg ctttgtaaca gcattaccat 13440
gcccccaggt ggaagctata tcttaaaggg cttgaaaatc cattcaagac agccgctaaa 13500
gatagctttt gactccctca cagaagattt ttcctcagct atgatatggg gaatgggtga 13560
gcagatggga gaagtaggaa gaagaggaga gaatgcttct tgggggtttg gaggggtgtt 13620
cagcatagtt ccacaatcaa accagcagga gagcagaact gtgaggcaac tctggggagg 13680
agttgaggct ctaggggaag tctcctgtag agcacaagca ggaaacatcc ggcctatagc 13740
agcattaaga agggctaatg tgtctcagga gggaaggatg ccatcaccat agaacctcta 13800
aatatgggca cagtaggatc ccagaaaagc agtgtttcgg ggaggatgcg ttctgcccaa 13860
aacatgtctg ttaaggttat tttgtagcac atggagcgct gatttgacct caagtttttg 13920
ttttttaaca ggtggaaagg caagtttaat ctacaatttt agtcgccacc aatacactct 13980
cttagagctt ttcatgacac gtctcataaa gaaatgctga tggccgggag cggtggctca 14040
cgcctgtaat cccagcactt tgggaggcca aggcgggcag attacgagat caggagatcc 14100
agagcatcct ggctaacacg gtgaaacccc gtctctacta aaaatacaaa aaattagccc 14160
ggcgtggtgg caggcgccta tagttccagc tactcgggag gctgaggcag gagaatggcg 14220
tgaacctggg aggtggagcg ggcagtgagc caagattgca ccactgcact ccaacctggg 14280
cgacagagcg agactctctc tcaaaaaaaa aaaaagaaaa aaagaaaaag aaaagaaaag 14340
aaaaaaaaag aaatgctgac gtttgccaag aggttcctga gttttggtca tactacagca 14400
cttgcaggca gtgtcactgc attcacatat aatgataata acgatattca cacatattaa 14460
gcacttattt atgctaggta tttttccaag ggatttacac atattaactc atttagattt 14520
tcacaacaac ctaatgaggt agctagtata cacatcttta tttcacagat gaggaaactg 14580
aagcatagag aggcaaaata aaccagccaa ggtcacatag ctaaccaagt ggtggagctg 14640
ggatttgtct aaaagtctgg tttcagaacc cttgtgctta atcctatact atactgttgg 14700
gtgtatcaac tgtatgctaa acagttgcct gtctggagcc aggacttcca gactttcagt 14760
ctgcacatat ggagccatac cactgacaag tatgtccaaa acttctttga tcctaagaat 14820
tacctggaca attgcaaaat atatagattc ccacaccctg gctcagatgt actcacaatc 14880
aggcaagttt ttaaaaccca ggtttagtgg gtttagtgag cactaccagc cagccctgag 14940
cattaggaaa ttgaagtttt tgtcctgatt ttgcttctgt ctctcagact ctgagcaatt 15000
tcactcttca attccctgct tgctctactg tctgcctgtc acttaacgga atgttacaag 15060
aatacataca atttttcccc ctcataaggg acacctgttg cttcaaaaac acggtatcct 15120
cataaaatga tatgcatgta gtaacaggtg tattttcttg cacttctttt gttttgtttt 15180
gttttttgct ttgctttcct tgaagcacaa acctaagccc ctcatccaga cctagccttc 15240
agctgtcctc caggtgacac gcatacacac cccaaaccag gctgcattct gaccgacctt 15300
agctctctcc ctctgggagc tctgatcggc tctcagttca gcccaacaat gagaaacttt 15360
tttctcgtct ccctcagggg agccttcacg tttatccaat tcattctctt gcaacccaac 15420
tctccagaaa gaaaaggggg gaaaatccca ccccgaagag acggtcttca ggtctgagga 15480
cgttacttag caacggcaca aagaccagtg agcaaaggga gacctgagga gaaaactctt 15540
gggtggggag acagagccag tttgaaaact ccatttcatc cagagaaaaa caaggaaaac 15600
acaaacagaa tcaatcccaa gtaacaagcg gggcttctcc ccagcgcagg tcatctctta 15660
ctccctgcat ctcaactcct tcaaaccccc agtgaccaag tccgcccccg cctggtttcg 15720
cccatggccc gagtgccctc cccttgcccc ggcctgaccc acacaggctt ggacttaggg 15780
gccccccacc ctccccaggc acccaccgtt ctcagacgcg ctgggacctt cgcagtccga 15840
gattaactgt tggggtttcc gctgctttcg ccgagacatt cccgggtaga gagttgggag 15900
agggaggggc aacgctcact tggtcttaac cggggtgacc tggtctcgtc tccccttgg 15960
gtccgaagcc aattgatgcc tctcccccag cgcaaatcac tgtgaagcag agatgttctt 16020
ctttcccaga gacacagact ctctctctct ctctgattct ctgttcttga ctctctctct 16080

<210> 3
<211> 1002
<212> PRT

<213> Mus musculus

<400> 3

```
Met Ala Gln Glu Thr Gly Ser Ser Ser Arg Leu Gly Gly Pro Cys Gly
 1               5                  10                  15
Glu Pro Ala Glu Arg Gly Gly Asp Ala Ser Glu Glu His His Pro Gln
            20                  25                  30
Val Cys Ala Lys Cys Cys Ala Gln Phe Ser Asp Pro Thr Glu Phe Leu
        35                  40                  45
Ala His Gln Asn Ser Cys Cys Thr Asp Pro Pro Val Met Val Ile Ile
    50                  55                  60
Gly Gly Gln Glu Asn Pro Ser Asn Ser Ser Ala Ser Ser Ala Pro Arg
65                  70                  75                  80
Pro Glu Gly His Ser Arg Ser Gln Val Met Asp Thr Glu His Ser Asn
                85                  90                  95
Pro Pro Asp Ser Gly Ser Ser Gly Pro Pro Asp Pro Thr Trp Gly Pro
            100                 105                 110
Glu Arg Arg Gly Glu Glu Ser Ser Gly Gln Phe Leu Val Ala Ala Thr
        115                 120                 125
Gly Thr Ala Ala Gly Gly Gly Gly Gly Leu Ile Leu Ala Ser Pro Lys
    130                 135                 140
Leu Gly Ala Thr Pro Leu Pro Pro Glu Ser Thr Pro Ala Pro Pro Pro
145                 150                 155                 160
Pro Pro Pro Pro Pro Pro Pro Pro Gly Val Gly Ser Gly His Leu Asn
                165                 170                 175
Ile Pro Leu Ile Leu Glu Glu Leu Arg Val Leu Gln Gln Arg Gln Ile
            180                 185                 190
His Gln Met Gln Met Thr Glu Gln Ile Cys Arg Gln Val Leu Leu Leu
        195                 200                 205
Gly Ser Leu Gly Gln Thr Val Gly Ala Pro Ala Ser Pro Ser Glu Leu
    210                 215                 220
Pro Gly Thr Gly Ala Ala Ser Ser Thr Lys Pro Leu Leu Pro Leu Phe
225                 230                 235                 240
Ser Pro Ile Lys Pro Ala Gln Thr Gly Lys Thr Thr Ala Ser Ser Ser
                245                 250                 255
Ser Ser Ser Ser Ser Ser Gly Ala Glu Pro Pro Lys Gln Ala Phe Phe
            260                 265                 270
His Leu Tyr His Pro Leu Gly Ser Gln His Pro Phe Ser Val Gly Gly
        275                 280                 285
Val Gly Arg Ser His Lys Pro Thr Pro Ala Pro Ser Pro Ala Leu Pro
    290                 295                 300
Gly Ser Thr Asp Gln Leu Ile Ala Ser Pro His Leu Ala Phe Pro Gly
305                 310                 315                 320
Thr Thr Gly Leu Leu Ala Ala Gln Cys Leu Gly Ala Ala Arg Gly Leu
                325                 330                 335
Glu Ala Ala Ala Ser Pro Gly Leu Leu Lys Pro Lys Asn Gly Ser Gly
            340                 345                 350
Glu Leu Gly Tyr Gly Glu Val Ile Ser Ser Leu Glu Lys Pro Gly Gly
        355                 360                 365
Arg His Lys Cys Arg Phe Cys Ala Lys Val Phe Gly Ser Asp Ser Ala
    370                 375                 380
Leu Gln Ile His Leu Arg Ser His Thr Gly Glu Arg Pro Tyr Lys Cys
385                 390                 395                 400
Asn Val Cys Gly Asn Arg Phe Thr Thr Arg Gly Asn Leu Lys Val His
                405                 410                 415
```

Phe His Arg His Arg Glu Lys Tyr Pro His Val Gln Met Asn Pro His
420 425 430
Pro Val Pro Glu His Leu Asp Tyr Val Ile Thr Ser Ser Gly Leu Pro
435 440 445
Tyr Gly Met Ser Val Pro Pro Glu Lys Ala Glu Glu Glu Ala Gly Thr
450 455 460
Pro Gly Gly Gly Val Glu Arg Lys Pro Leu Val Ala Ser Thr Thr Ala
465 470 475 480
Leu Ser Ala Thr Glu Ser Leu Thr Leu Leu Ser Thr Gly Thr Ser Thr
485 490 495
Ala Val Ala Pro Gly Leu Pro Thr Phe Asn Lys Phe Val Leu Met Lys
500 505 510
Ala Val Glu Pro Lys Ser Lys Ala Asp Glu Asn Thr Pro Pro Gly Ser
515 520 525
Glu Gly Ser Ala Ile Ala Gly Val Ala Asp Ser Gly Ser Ala Thr Arg
530 535 540
Met Gln Leu Ser Lys Leu Val Thr Ser Leu Pro Ser Trp Ala Leu Leu
545 550 555 560
Thr Asn His Leu Lys Ser Thr Gly Ser Phe Pro Phe Pro Tyr Val Leu
565 570 575
Glu Pro Leu Gly Ala Ser Pro Ser Glu Thr Ser Lys Leu Gln Gln Leu
580 585 590
Val Glu Lys Ile Asp Arg Gln Gly Ala Val Ala Val Ala Ser Thr Ala
595 600 605
Ser Gly Ala Pro Thr Thr Ser Ala Pro Ala Pro Ser Ser Ser Ala Ser
610 615 620
Gly Pro Asn Gln Cys Val Ile Cys Leu Arg Val Leu Ser Cys Pro Arg
625 630 635 640
Ala Leu Arg Leu His Tyr Gly Gln His Gly Gly Glu Arg Pro Phe Lys
645 650 655
Cys Lys Val Cys Gly Arg Ala Phe Ser Thr Arg Gly Asn Leu Arg Ala
660 665 670
His Phe Val Gly His Lys Thr Ser Pro Ala Ala Arg Ala Gln Asn Ser
675 680 685
Cys Pro Ile Cys Gln Lys Lys Phe Thr Asn Ala Val Thr Leu Gln Gln
690 695 700
His Val Arg Met His Leu Gly Gly Gln Ile Pro Asn Gly Gly Ser Ala
705 710 715 720
Leu Ser Glu Gly Gly Gly Ala Ala Gln Glu Asn Ser Ser Glu Gln Ser
725 730 735
Thr Ala Ser Gly Pro Gly Ser Phe Pro Gln Pro Gln Ser Gln Gln Pro
740 745 750
Ser Pro Glu Glu Glu Met Ser Glu Glu Glu Glu Glu Asp Glu Glu Glu
755 760 765
Glu Glu Asp Val Thr Asp Glu Asp Ser Leu Ala Gly Arg Gly Ser Glu
770 775 780
Ser Gly Gly Glu Lys Ala Ile Ser Val Arg Gly Asp Ser Glu Glu Val
785 790 795 800
Ser Gly Ala Glu Glu Glu Val Ala Thr Ser Val Ala Ala Pro Thr Thr
805 810 815
Val Lys Glu Met Asp Ser Asn Glu Lys Ala Pro Gln His Thr Leu Pro
820 825 830
Pro Pro Pro Pro Pro Pro Asp Asn Leu Asp His Pro Gln Pro Met Glu
835 840 845
Gln Gly Thr Ser Asp Val Ser Gly Ala Met Glu Glu Glu Ala Lys Leu

```
    850                 855                 860
Glu Gly Ile Ser Ser Pro Met Ala Ala Leu Thr Gln Glu Gly Glu Gly
865                 870                 875                 880
Thr Ser Thr Pro Leu Val Glu Glu Leu Asn Leu Pro Glu Ala Met Lys
                885                 890                 895
Lys Asp Pro Gly Glu Ser Ser Gly Arg Lys Ala Cys Glu Val Cys Gly
            900                 905                 910
Gln Ser Phe Pro Thr Gln Thr Ala Leu Glu Glu His Gln Lys Thr His
        915                 920                 925
Pro Lys Asp Gly Pro Leu Phe Thr Cys Val Phe Cys Arg Gln Gly Phe
    930                 935                 940
Leu Asp Arg Ala Thr Leu Lys Lys His Met Leu Leu Ala His His Gln
945                 950                 955                 960
Val Pro Pro Phe Ala Pro His Gly Pro Gln Asn Ile Ala Thr Leu Ser
                965                 970                 975
Leu Val Pro Gly Cys Ser Ser Ser Ile Pro Ser Pro Gly Leu Ser Pro
            980                 985                 990
Phe Pro Arg Lys Asp Asp Pro Thr Met Pro
        995                 1000
```

<210> 4
<211> 4547
<212> DNA
<213> Mus musculus

<400> 4

```
atggcgcagg aaaccgggag cagctctcga ctcgggggac cctgcgggga gcctgcggag 60
cgcggaggtg atgctagcga ggaacaccac ccccaagtct gtgccaaatg ctgcgcacaa 120
ttttctgacc cgaccgaatt cctcgctcac cagaactcat gttgcactga cccaccggta 180
atggtgataa ttggaggcca ggagaatccc agcaactctt cagcctcctc tgcgccccga 240
ccagagggcc acagtaggtc ccaggtcatg gatacagagc acagcaatcc cccagattct 300
gggtcctctg ggcccccgga tcccacttgg gggccagagc ggaggggaga ggaatcttct 360
gggcaattcc tggtcgctgc cacaggtaca gcggctgggg gaggtggggg ccttatcttg 420
gccagtccca agctgggagc aaccccatta cctccagaat ccactcctgc acccccctcct 480
cccccaccac ccccctcccc tccaggtgta ggcagtggcc acttgaacat tcctctgatc 540
ttggaagagt tgcgggtgct gcagcagcgc cagattcacc agatgcagat gactgaacaa 600
atctgccgcc aggtgctgct acttggctcc ttggggcaga ccgtgggtgc ccctgccagt 660
ccctcagagc tacctgggac aggggctgcc tcttccacca agcccctcct gcctctcttc 720
agtcccatca agccagcgca aactggcaag acactggcat cttcctcttc gtcatcctcc 780
tcctctggag ctgaaccgcc taagcaggct ttcttccacc tttaccatcc actgggatca 840
cagcatcctt tctctgtagg aggggttggg cggagccaca aacccacccc tgccccttcc 900
cctgcgctgc caggcagtac ggatcagctg attgcttcac ctcatctggc attcccaggc 960
accactggac tcctggcagc tcagtgtctt ggggcagcaa ggggccttga ggctgctgcc 1020
tccccagggc tcctgaagcc aaagaacgga agtggtgaac tgggctatgg ggaagtgatc 1080
agttccttgg agaaacccgg tggaaggcac aaatgccgct tttgtgcaaa agtattcggc 1140
agtgacagcg ccctgcagat ccaccttcgt tcccacactg gtgagaggcc ctataagtgc 1200
aacgtctgtg gtaaccgttt cacaactcgg ggcaacctca aagtacattt tcaccggcat 1260
cgtgagaagt acccacatgt gcaaatgaat ccacatccag taccggagca cctagactac 1320
gtcatcacca gcagtgggct gccttacgga atgtctgtgc caccagagaa agcagaagag 1380
gaggcaggca caccaggcgg aggtgttgaa cgcaaacccc tagtggcctc caccacagca 1440
ctcagtgcca cagagagcct gacactgctc tccactggca caagcacagc agtggctcct 1500
gggctcccta ctttcaacaa gtttgtgctc atgaaggcag tggaacccaa gagtaaagcg 1560
gatgagaaca cgcccccagg gagtgagggc tccgccatcg ctggagtagc agacagtggc 1620
tcagcaaccc gaatgcagct aagtaagctg gtgacgtcac taccgagttg ggcactgctt 1680
```

actaatcact tgaagtcaac tggaagtttc cccttccctt atgtgctaga acccttgggg 1740
gcttcgcctt ctgagacctc aaagctgcag cagctagtag aaaagattga ccgccaagga 1800
gctgtggcgg tggcatctac tgcctcggga gctcccacca cttctgcccc tgcaccttcc 1860
tcctccgctt ctggacctaa ccagtgtgtg atctgtcttc gggtcctgag ctgccctcgg 1920
gctctacgcc tgcattatgg ccaacatgga ggtgagcggc ccttcaagtg taaagtgtgt 1980
ggccgagctt tctccacaag gggcaatttg cgcgcacatt tcgtgggtca caagaccagt 2040
ccagctgccc gggctcagaa ctcctgcccc atttgtcaga agaagttcac taatgctgtc 2100
actctgcagc aacatgttcg gatgcacctg gggggccaga tccccaatgg gggttccgca 2160
ctttctgaag gtgggggagc tgcccaggaa aacagctctg agcagtctac agcctctgga 2220
ccagggagtt tcccccagcc gcagtcccag cagccatctc cagaagagga gatgtctgag 2280
gaagaggaag aggatgagga agaggaggaa gacgtgacag atgaagattc cctagcagga 2340
agaggctctg agagtggggg agagaaggcc atatcagtac gaggtgactc agaagaggta 2400
tctggggcag aggaagaagt ggcaacatca gtagcagcac ccaccactgt gaaggagatg 2460
gacagtaatg agaaagcccc tcaacacact ctgccgccac ctccgccacc acccgacaac 2520
ctggatcatc cccaacccat ggagcaggga accagtgatg tttccggagc catggaggaa 2580
gaagccaaac tggagggaat ctcaagcccg atggcagccc tcacccaaga aggggagggc 2640
accagcaccc ctttggtgga agagctgaac ttaccggaag ccatgaagaa ggatccagga 2700
gagagcagcg gcaggaaggc ctgtgaagta tgtggccaga gctttcctac ccagacagct 2760
ctggaggagc atcagaagac ccatcccaag gatgggccac tcttcacttg tgtcttctgc 2820
aggcagggct tccttgaccg tgctaccctc aagaagcaca tgctgttggc tcaccaccag 2880
gtaccgccct ttgcacccca tggccctcag aatattgcta ctctttcctt ggtccctggc 2940
tgttcctcct ccatcccttc tccagggctc tccccattcc ctcgaaaaga tgaccccacc 3000
atgccatgag cctgctttct gtacctggtc ctctatgacc cagagagcag aaacctgaga 3060
gcttcataga ggaactccaa gatttactca ccctcctctt gtcctttctc aagtcctgac 3120
atgatgtttc tagtggcttc ttctctagtc cctgagcttg acaattgcct ttgaaagaga 3180
atgtcccctt aagaaatttt tatcaccttt ttgttctgtg taactaaggg aaacaaattc 3240
cctatagctt ttacattctc aaggggggagc tctctcctct tctccctttc cctttggcag 3300
gtatactaga acccccatcc ttggagtggc agccttggtc caaggggctg gcaactgtcc 3360
atggaaggcc cagcgttact ccttggtgat cttgaccacc ctgcaagact ttctagggcc 3420
gggaccttct tgagaagctt gtaaggggtg gtaggtttct ttctgcaacc actacccagt 3480
tttccactga gccctggagt tctggaccta cctgcattgc cactcgggcc ctagtaccat 3540
cattgctgtg aaagcccagg aactgtgttt cacaaggtga ctccagtgac atgatccaga 3600
gaggcaaaga acatagcctc cggaagttga ggctgtgccc aacaagcaca ccggaagaaa 3660
gaagaaacta taacttcttt ctccttcccc cctgctccag agagtgctgg caataaagat 3720
attctagcaa ttggtgactc accctagaag gtagggacaa gtgaaggact gggacccttt 3780
ttgcagtatg ttccttgact cgccacattg aggcaaagat agtggctggt caagatgcca 3840
ggactactcc agcttcccat catgtcctct caaccaacaa gcaggtttcc taccaagagg 3900
tctctcgtgt gatagtttag ggagtatgaa gtttctaact ctaaagaatc ctgttggtga 3960
ggatgattat ttaagcaatg atggggagtt gagggttgtt gctaaaacag gcattgctgg 4020
gaatctattt gatgaagaac aggacttgat gtaaggggac tcgatgttca gctcttgtga 4080
gtatgaacgt tttctttgag ctaatggtga tgtggtatgc agaggtacca ggggccatgg 4140
gggtgtgtgt gcttcctgtc actagaatgt ttttagtttt agatgactcc ctattttatt 4200
ccctcacccc ttgtatttcc cttgctgtct tctcaaaacc cctttcctcc cccagttttg 4260
cctgaccatg ggccagagct tatgtcttat ttttttttcta gaagttgaga gacagagctt 4320
caagtggttt ccccccgtct ctgtcttgta gtgagatgta gtatttactc ttaacatagg 4380
atcctgtgga acaggtgttc tgagaagact gaattttgct gttagctgtt gtcaatgatg 4440
attctctaaa gtagtgggct ccagagctcc ctaacacagt gaaatgtgta agagccgaga 4500
ggggagatac tagaattttt tccttcatca ttaaaggtgt tttggct 4547

<210> 5
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> derived from human Sal2 gene

<400> 5
ccacaaccat ggcgaatccg ag 22

<210> 6
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from human Sal2 gene

<400> 6
ggtgatggaa ggcgaacagc cagg 24

<210> 7
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from Human Sal2 gene

<400> 7
cttgttaatt agagcctcgg tatacc 26

<210> 8
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from Human Sal2 gene

<400> 8
gcacggagga cccagaatct gg 22

<210> 9
<211> 63
<212> DNA
<213> Polyoma virus

<400> 9
gatatacttt gtaatgtgca agaaggcgac gacccttga aggacatatg tgaatatagc 60
tga 63

<210> 10
<211> 20
<212> PRT
<213> Polyoma virus

<400> 10
Asp Ile Leu Cys Asn Val Gln Glu Gly Asp Asp Pro Leu Lys Asp Ile
1 5 10 15

Cys Glu Tyr Ser
20

<210> 11
<211> 19
<212> PRT
<213> TMD25 mutant Polyoma virus

<400> 11
Asp Ile Leu Cys Asn Val Gln Glu Asp Phe Val Met Cys Lys Lys Ala
1 5 10 15
Thr Thr Pro

<210> 12
<211> 60
<212> DNA
<213> TMD25 mutant Polyoma virus

<400> 12
gatatacttt gtaatgtgca agaagacttt gtaatgtgca agaaggcgac gacccttga 60

<210> 13
<211> 16
<212> PRT
<213> Polyoma virus

<400> 13
Asn Val Gln Glu Gly Asp Asp Pro Leu Lys Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 14
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 14
Asn Val Gln Glu Gly Asp Asp Pro Leu Lys Asp Ile Cys Glu
1 5 10

<210> 15
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 15
Asn Val Gln Glu Gly Asp Asp Pro Leu Lys
1 5 10

<210> 16
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 16
Asn Val Gln Glu Gly Asp Asp
1 5

<210> 17
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 17
Asn Val Gln Glu
1

<210> 18
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 18
Asn Val Gln Glu Gly Asp Asp Leu Lys Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 19
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 19
Asn Val Gln Glu Gly Asp Asp Pro Lys Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 20
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 20
Asn Val Gln Glu Gly Asp Asp Pro Leu Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 21
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 21
Asn Val Gln Glu Gly Asp Asp Asp Ile Cys Glu Tyr Ser
1 5 10

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/04596 A3

(51) International Patent Classification[7]: C12Q 1/68, C12N 5/00, C07H 21/02, 21/04

(21) International Application Number: PCT/US01/21354

(22) International Filing Date: 5 July 2001 (05.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/216,723	7 July 2000 (07.07.2000)	US
09/812,633	19 March 2001 (19.03.2001)	US
09/812,471	19 March 2001 (19.03.2001)	US

(71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). **THE BRIGHAM AND WOMEN'S HOSPITAL, INC.** [US/US]; 75 Francis Street, Boston, MA 02115 (US).

(72) Inventors: **BENJAMIN, Thomas, L.**; 595 Putnam Avenue, Cambridge, MA 02139 (US). **LI, Dawei**; 16 Copenger Street, Apartment #1, Boston, MA 02120 (US). **MOK, Samuel, C.**; 125 Pleasant Street, #607, Brookline, MA 02446 (US). **CRAMER, Daniel, W.**; 151 Shaw Road, Chestnut Hill, MA 02167 (US). **MA, Yupo**; 24 West Street, Sharon, MA 02067 (US).

(74) Agent: **BIEKER-BRADY, Kristina**; Clark & Elbing LLP, 176 Federal Street, Boston, MA 0211-2214 (US).

(81) Designated States *(national)*: CA, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(88) Date of publication of the international search report: 18 April 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/04596 A3

(54) Title: DIAGNOSING AND TREATING CANCER CELLS USING MUTANT VIRUSES

(57) Abstract: The invention provides methods for the identification of genes and their encoded proteins involved in the susceptibility to proliferative disorders, including cancer, using a tumor host range mutant virus (T-HR mutant). In addition, the invention provides methods for the diagnosis of abnormally proliferating cells in a subject, using a T-HR mutant. The invention also features, T-HR mutants that can be used to kill cancer cells such as ones carrying a Sal2 alteration. Furthermore, the invention features the analysis of Sal2 nucleic acids and proteins for diagnosing and treating patients having proliferative disorders, including cancer, involving mutations in a *Sal2* gene and encoded protein. Also encompassed by the invention are transgenic and knockout mice including Sal2 nucleic acids or proteins and mutants thereof.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/21354

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12Q 1/68; C12N 5/00; C07H 21/02, 21/04
US CL : 435/6, 325; 536/23.1, 24.5; 935/1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 325; 536/23.1, 24.5; 935/1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST, MEDLINE, BIOSIS, CAPLUS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FREUND et al. Host range and cell cycle activation properties of polyomavirus large T-antigen mutants defective in pRB binding. J Virology. November 1994, vol. 68, no. 11, pages 7227-7234, see entire document.	1, 2, 7-11
A	US 6,214,544 B1 (FISHER) 10 APRIL 2001, SEE ABSTRACT AND CLAIMS.	1-11

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 November 2001 (15.11.2001)	13 FEB 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Q. JANICE LI Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/21354

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-11

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/21354

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional examination fees must be paid.

Group I, claims 1-11, drawn to a method identifying a cellular protein involved in the susceptibility to proliferative disease.

Group II, claim 12, drawn to a tumor host range virus.

Group III, claims 13-19, drawn to a method of determining gene alteration comprising determining whether a cell can act as a permissive host for the propagation of a T-HR mutant.

Group IV, claims 20-30, drawn to a method for cancer cell killing.

Group V, claims 31-42, drawn to a method for risk assessment of a proliferative disease comprising detecting alteration of a *Sal2* nucleic acid in a mammal.

Group VI, claims 43-48, drawn to a method for risk assessment of a proliferative disease comprising detecting alteration of a *Sal2* protein in a mammal.

Group VII, claims 49-52, 54, 59, 60, drawn to a knockout mouse comprising a mutation in a genomic mSal2 gene.

Group VIII, claims 53, 55-58, 61, drawn to a transgenic mouse whose genome comprising a nucleic acid construct including a Sal2 gene.

Group IX, claims 62-69, 72, 74, 75, drawn to a method of identifying a compound which alters cell proliferation by measuring cell proliferation of two different abnormal test cells.

Group X, claims 70, 71, 73, 76, drawn to a method of identifying a compound which alters cell proliferation by measuring cellular *Sal2* levels.

The inventions listed as Groups I-X do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I is drawn to a method of identifying a cellular protein, whereas group II is drawn to a tumor host range virus. Thus, they lack the same special technical feature.

The inventions listed as Groups III-VI, IX, X, and I do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I, III-VI, IX, X are drawn to different methods, such as for identifying a cellular protein, assessing a gene alteration or cell killing. The different methods using different test criteria, different starting materials, have different method steps, different mode of operation, and different technical considerations. Thus, they lack the same special technical feature.

The inventions listed as Groups VI and V do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups VI differs from group V in that they are different methods for risk assessment, the different method using different test criteria, i.e. a Sal2 nucleic acid or a Sal2 protein, the method steps and mode of operation for identifying a nucleic acid differs from that of a protein. Thus, they lack the same special technical feature.

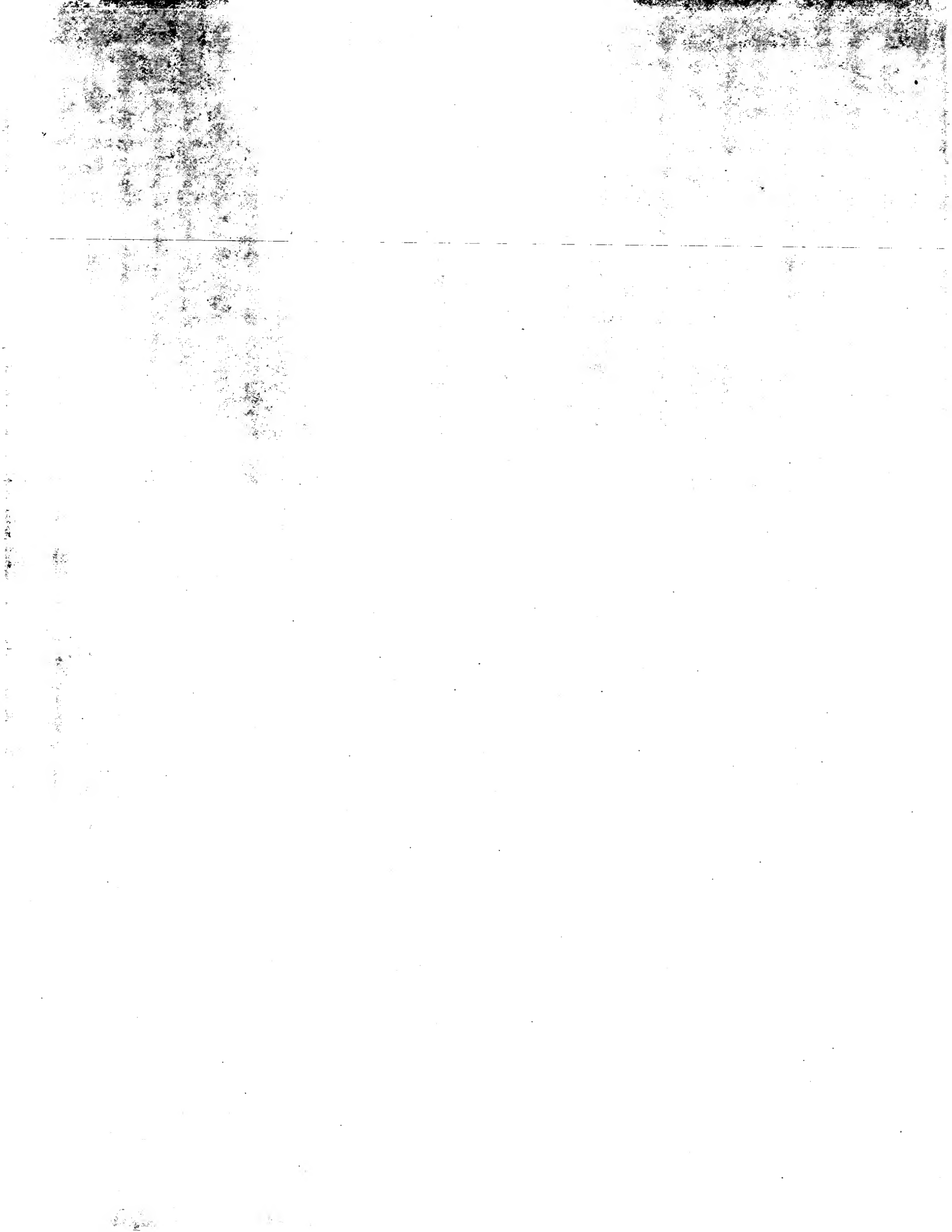
The inventions listed as Groups X and IX do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups X differs from group IX in that they are different methods for identifying a compound, the different method using different test criteria, i.e. state of cell proliferation or levels of Sal 2 expression, the method steps and mode of operation for the two differs significantly that they lack the same special technical feature.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/21354

The inventions listed as Groups VIII and VII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups VIII differs from group VII in that they are different products, a Sal2 transgene mouse and a Sal2 gene knockout mouse differs in their genomic sequence structures and phenotype . Thus, they lack the same special technical feature.



(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 January 2002 (17.01.2002)

PCT

(10) International Publication Number
WO 02/004596 A3

(51) International Patent Classification[7]: C12Q 1/68, C12N 5/00, C07H 21/02, 21/04

(21) International Application Number: PCT/US01/21354

(22) International Filing Date: 5 July 2001 (05.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/216,723	7 July 2000 (07.07.2000)	US
09/812,633	19 March 2001 (19.03.2001)	US
09/812,471	19 March 2001 (19.03.2001)	US

(71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). **THE BRIGHAM AND WOMEN'S HOSPITAL, INC.** [US/US]; 75 Francis Street, Boston, MA 02115 (US).

(72) Inventors: **BENJAMIN, Thomas, L.**; 595 Putnam Avenue, Cambridge, MA 02139 (US). **LI, Dawei**; 16 Copenger Street, Apartment #1, Boston, MA 02120 (US). **MOK, Samuel, C.**; 125 Pleasant Street, #607, Brookline, MA 02446 (US). **CRAMER, Daniel, W.**; 151 Shaw Road, Chestnut Hill, MA 02167 (US). **MA, Yupo**; 24 West Street, Sharon, MA 02067 (US).

(74) Agent: **BIEKER-BRADY, Kristina**; Clark & Elbing LLP, 176 Federal Street, Boston, MA 0211-2214 (US).

(81) Designated States *(national)*: CA, JP.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*

(88) Date of publication of the international search report:
18 April 2002

(48) Date of publication of this corrected version:
6 February 2003

(15) Information about Correction:
see PCT Gazette No. 06/2003 of 6 February 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DIAGNOSING AND TREATING CANCER CELLS USING MUTANT VIRUSES

(57) Abstract: The invention provides methods for the identification of genes and their encoded proteins involved in the susceptibility to proliferative disorders, including cancer, using a tumor host range mutant virus (T-HR mutant). In addition, the invention provides methods for the diagnosis of abnormally proliferating cells in a subject, using a T-HR mutant. The invention also features, T-HR mutants that can be used to kill cancer cells such as ones carrying a Sal2 alteration. Furthermore, the invention features the analysis of Sal2 nucleic acids and proteins for diagnosing and treating patients having proliferative disorders, including cancer, involving mutations in a *Sal2* gene and encoded protein. Also encompassed by the invention are transgenic and knockout mice including Sal2 nucleic acids or proteins and mutants thereof.

WO 02/004596 A3

DIAGNOSING AND TREATING CANCER CELLS USING MUTANT VIRUSES

Statement as to Federally Sponsored Research

The present research was supported by a grant from the National Cancer Institute (Number R35 CA44343). The U.S. government has certain rights to this invention.

Field of the Invention

The field of the invention is regulation of cellular proliferation.

Background of the Invention

Transforming genes of DNA tumor viruses perform essential functions in virus growth, acting largely as proto-oncogene activators or tumor suppressor gene inactivators. The isolation and characterization of mutant viruses that are able to propagate in cells containing a mutation in known proto-oncogene or tumor suppressor genes has been useful in identifying and studying the viral equivalents or interactors of these genes. The transforming gene of the highly oncogenic murine polyoma virus was identified through studies of host range mutants isolated using polyoma transformed 3T3 cells as the permissive host and normal 3T3 cells as the non-permissive host. This approach requires expression a known viral protein by the permissive host, since it is based on the idea of complementation between cell-associated wild-type viral genes and an infecting virus mutant. In addition to its use with polyoma virus, the complementation approach has also been successfully used with other oncogenic DNA viruses, e.g., with 293 cells expressing adenovirus E1A/E1B genes and COS cells expressing the SV40 large T antigen. Complementing cell lines have also been used in other systems to propagate specifically defective virus mutants for vaccine development and other purposes. However, by design, these types of

systems rely on permissive hosts constructed with known gain-of-function mutations and are only applicable to mutants in known viral genes, as well as to viruses with known mutations, since the host cell must express a functional version of the mutant viral protein.

The use of mutant adenoviruses unable to inactivate p53 or the retinoblastoma protein (pRb) to kill cancer cells lacking one of these proteins has been previously described (Patent Nos. U.S. 5,677,178 and WO 94/18992). It was well known prior to these observations that these two genes are mutated in a variety of cancers.

While a number of genes are known to be involved in the progression towards cancer, there is a significant need for the development of a general, unbiased method for identifying new genes involved in the pre-disposition for, or progression of cancer or other proliferative disorders. Furthermore, methods for diagnosing and treating patients with mutations in known as well as newly identified genes would greatly aid in the management of cancer.

Summary of the Invention

The invention features novel tumor host range viruses, for identifying mammalian cancer susceptibility genes, such as tumor suppressor genes and proto-oncogenes, and methods for diagnosing and treating patients having proliferative disorders, such as cancers, involving mutations in such genes. Furthermore, the invention features the use of *Sal2* nucleic acids and proteins in methods of identifying a mammal having, or at risk of acquiring, a proliferative disease.

The tumor host range mutant viruses (T-HR mutants) used in the methods of the invention contain mutations that prevent the virus from propagating in normal cells. These viruses are, however, able to propagate in abnormally proliferating cells because of genetic changes that are present in these cells, such as the inactivation of tumor suppressor genes or the activation of proto-oncogenes. A T-HR mutant that infects a normal cell is unable to propagate in such a cell because it is unable to inactivate a tumor suppressor gene or to

activate a proto-oncogene due to a mutation in the viral genome. In contrast, if this T-HR mutant infects an abnormally proliferating cell that already has a tumor suppressor gene inactivated, this virus is able to propagate. Likewise, if such a T-HR mutant infects an abnormally proliferating cell that contains an activated proto-oncogene, the virus is also able to propagate.

The Tumor Host Range Mutant System

Since a T-HR mutant is unable to propagate in normal cells, but is able to propagate in abnormally proliferating cells, the first aspect of the invention features a method of using T-HR mutants to identify a cellular protein that is involved in the susceptibility to cancer and other proliferative disorders. This method involves: (a) infecting a normal cell and an abnormally proliferating cell with a collection of uncharacterized mutant viruses; (b) identifying a mutant virus from the collection that can grow in an abnormally proliferating cell and can not grow in a normal cell (i.e., a T-HR mutant); (c) identifying the mutated viral gene or mutated protein in the virus, where this mutation allows the virus to grow on the abnormally proliferating cell; and (d) screening to identify the cellular proteins which interact with the wild-type viral protein, but not with the mutated protein.

In a preferred embodiment of the above aspect of the invention, the abnormally proliferating cell infected with the collection of uncharacterized mutant viruses is also uncharacterized. In an additional preferred embodiment, the cellular and viral proteins can be identified by, for example, using an assay that detects protein-protein interactions (e.g., a GST-pull-down assay). These proteins may be, for example, tumor suppressor proteins or proto-oncogene products; however the retinoblastoma tumor suppressor protein and the gene encoding this protein are specifically excluded from this and all other aspects of the invention. In another preferred embodiment, the method of this aspect is used to isolate a mutant virus (i.e., a T-HR mutant).

Preferred viruses with a mammalian, preferably human, host range used in this and other aspects of the invention include, for example, simian virus 40,

human polyoma virus, parnovirus, papilloma virus, herpes virus, and primate adenoviruses.

The second aspect of the invention features a method of determining the presence or absence of an alteration in the genetic material of a cell that involves determining whether such a cell can act as a permissive host for the growth of a characterized T-HR mutant, where the T-HR mutant is capable of propagating in an abnormally proliferating cell and not capable of propagating in a normal cell. The retinoblastoma and p53 genes are specifically excluded from this aspect of the invention.

In a preferred embodiment of the above aspect of the invention, the alteration of the genetic material to be tested for in the cell indicates that the organism carrying this alteration is at an increased risk of developing a proliferative disease. Preferably, this genetic alteration is in a tumor suppressor gene or in a proto-oncogene. In another preferred embodiment, the T-HR mutant has been characterized as being complemented by a mutation in a specific tumor suppressor or proto-oncogene. In an additional preferred embodiment of the above aspects of the invention, the cells used in the methods of the invention are from a mammal, for example, a human.

In another aspect, the invention features a method of killing a cell with a proliferative disease that involves: (i) contacting a cell with a proliferative disease, for example, a mammalian cell, with a T-HR mutant; and (ii) allowing the T-HR mutant to lyse this cell. In a preferred embodiment, the TH-R mutant may be a T-HR mutant specific for a cell carrying a *Sal2* mutation, for example, the TMD-25 T-HR mutant virus. In an additional preferred embodiment of this aspect, the mammalian cell is from a human. The mammalian cell may also be in a mammal, for example a human, with a proliferative disorder. In a further embodiment, the T-HR mutant may be administered, for example, in a pharmaceutically acceptable carrier. In addition, the T-HR mutant may be administered, for example, by parenteral, intravenous, intraperitoneal, intramuscular, subcutaneous, or subdermal injection. The T-HR mutant, however, may also be administered orally, nasally, topically, or as an aerosol.

The Use of Sal2 as a Diagnostic and Treatment Tool

A further aspect of the invention features a method of identifying a mammal having, or at increased risk of acquiring, a proliferative disease. This method includes determining whether there is a proliferative disease-associated alteration in a *Sal2* nucleic acid of the mammal. An example of a proliferative disease-associated Sal2 alteration is the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1. In one embodiment, the method is used to identify a mammal, preferably a human, having a proliferative disease, while in another embodiment, the method is used to identify a mammal at increased risk of acquiring a proliferative disease.

In another embodiment of this aspect of the invention, determining whether the mammal has or is at increased risk of acquiring a proliferative disease is done by, for example, polymerase chain reaction (PCR) amplification, single nucleotide polymorphism (SNP) determination, restriction fragment length polymorphism (RFLP) analysis, hybridization analysis, or mismatch detection analysis.

In addition, identifying the alteration may also involve: (i) contacting a first nucleic acid probe which is specific for binding to the human *Sal2* nucleic acid containing the alteration with a nucleic acid from a cell from the mammal under conditions which allow the first nucleic acid probe to anneal to complementary sequences in the cell; and (ii) detecting duplex formation between the first nucleic acid probe and the complementary sequences. The nucleic acid probe of step (i), which is, for example, at least 12 contiguous nucleotides in length, may be derived from the human *Sal2* nucleic acid containing a proliferative disease-associated alteration. The cell may be from a physiological sample, which may contain, for example, mRNA or the nucleic acid probe of step (i) annealed to the mRNA. Furthermore, another embodiment of this aspect includes a second nucleic acid probe, where the first and second nucleic acid probes are PCR primers, and where the human *Sal2* nucleic acid or a fragment is amplified using PCR between steps (i) and (ii).

In another embodiment of this aspect, the cell may be selected from a physiological sample, for example, containing abnormally proliferating or normal tissue, and may be from human tissue, blood, ovarian tissue, bladder tissue, colon tissue, and cells grown in culture.

An additional aspect of the invention features a method of identifying a mammal having, or at increased risk of acquiring, a proliferative disease involving, determining whether there is an alteration in a Sal2 protein of the mammal. Preferably the method is used to identify a mammal, such as a human, having a proliferative disease, or at increased risk of acquiring a proliferative disease. In preferred embodiments of this method an antibody specific for either the human, or for a proliferative disease-associated mutant Sal2 protein is used.

A further aspect of the invention encompasses a knockout mouse featuring a knockout mutation in a genomic *mSal2* gene. This knockout mouse may also contain, for example, a nucleic acid construct including a mutant *Sal2* gene and this mutant *Sal2* gene may be conditionally expressed. In a preferred embodiment, the mutant *Sal2* gene, for example a human *Sal2* gene, encodes a protein that contains a substitution of a Cys for the Ser at position 73 of SEQ ID NO:1. However, the Sal2 protein may also be wild-type.

An additional aspect of the invention features a transgenic mouse whose genome includes a nucleic acid construct that contains a *Sal2* nucleic acid, which is operably linked to transcriptional regulatory elements and encodes a Sal2 protein, e.g., a mutant Sal2 protein. The mutant Sal2 protein may also be a human Sal2 protein, for example, one that has a modification of function. Furthermore, the transgenic mouse may contain a mouse Sal2 protein, e.g., the protein of SEQ ID NO:3. This mouse Sal2 protein may be mutant, such as a mouse Sal2 protein containing the substitution of a Cys for the Ser at position 73 of SEQ ID NO:3.

In preferred embodiments of this aspect, the transcriptional regulatory elements include a promoter that is a tissue-specific promoter, such that the nucleic acid is expressed, and the protein is produced at detectable levels, in cells selected from the group consisting of ovarian, bladder, and colon cells.

However, the transcriptional regulatory element may also include the wild-type *Sal2* promoter.

In a further embodiment of this aspect, the transgenic mouse develops ovarian tumors, and these tumors may metastasize. The invention also includes a cell line, such as an ovarian cell line, derived from cells isolated from the transgenic mouse.

An additional aspect, the invention encompasses a method of identifying a compound which alters cell proliferation, the method involving: a) contacting a first cell with a test compound, and b) measuring whether the test compound alters proliferation in the first cell, relative to a second cell not contacted with the test compound, wherein the first and second cells have a proliferative disease-associated alteration in a *Sal2* nucleic acid. In a preferred embodiment of this aspect, the ability of the test compound to alter proliferation is determined by measuring the ability of a virus, for example, a T-HR mutant virus, to propagate in the first cell contacted with the test compound, relative to the second cell not contacted with the test compound. In addition, the first and second cells may be mammalian cells, for example, human cells. Furthermore, these cells may be ovarian cells. The cells may also be in the same mammal or in different mammals and the mammal may be a transgenic mouse or a knockout mouse containing a knockout mutation in a genomic *mSal2* gene.

A final aspect of the invention features a method of identifying a compound which alters cell proliferation, the method involving: a) exposing a cell or a cell extract to a test compound, and b) measuring whether the test compound alters Sal2 levels, for example, Sal2 protein or nucleic acid levels, relative to Sal2 levels in a cell or cell extract not exposed to the test compound. In a preferred embodiment of this aspect, the cell has a proliferative disease-associated alteration in a *Sal2* nucleic acid or the extract is from a cell having a proliferative disease-associated alteration in a *Sal2* nucleic acid. This cell or cell extract may be mammalian, e.g., human. Furthermore, the cell may be in a mammal, for example, a transgenic mouse whose genome includes a nucleic acid construct containing a *Sal2* nucleic acid, which is operably linked to

transcriptional regulatory elements and encodes a Sal2 protein, or a knockout mouse comprising a knockout mutation in a genomic *mSal2* gene. In another embodiment, the exposing in step a) of this aspect is done with a cell and this cell is an ovarian cell.

Definitions

"Tumor host range mutant virus (T-HR mutant)," as used herein, refers to a virus that has a reduced ability to replicate and disseminate in a normal cell, relative to the replication of a wild-type virus in the same type of cell, but is able to replicate and disseminate in a cell having abnormal proliferation. The abnormally proliferating cell may, for example, have one or more mutations in a gene or genes involved in the regulation of cell growth, of the cell cycle, or of programmed cell death (e.g., apoptosis). These genes include, for example, tumor suppressor genes and proto-oncogenes, but any cellular gene that a virus must inactive or activate in order to grow is also included. Adenoviruses having mutations in the p53 and retinoblastoma genes are specifically excluded.

Reference herein to a "collection of uncharacterized mutant viruses" refers to a sample of viruses where at least one of the viruses, in a collection of at least 1000 viruses, (e.g., 0.1%) carries at least one mutation in at least one of the genes of the viral genome. Preferably, at least 10%, 25%, 30%, or 50% of the viruses in this collection carry at least one mutation in at least one of the genes in the viral genome. In addition, such mutations preferably inactivate viral proteins that are necessary for transforming a host cell into a cancer cell. The types of mutations that may be present in the viral genes include, for example, point mutations, deletions, insertions, duplications, and inversions. Furthermore, the mutations may result in modification of function, such as a partial or a complete loss-of-function of the viral gene. Preferably the virus has a mammalian host range (e.g., rodent or primate), most preferably a human host range. Viruses that may be used in such a collection include, for example, simian virus 40, human polyoma virus, parnovirus, papilloma virus, herpes virus, and primate adenoviruses. However, any virus that needs to overcome a cell cycle checkpoint

or affect a signal transduction pathway in order to propagate may be used in this collection.

"T-HR mutant specific for a Sal2 mutation," as used herein, refers to a TH-R mutant virus that is able to propagate in a cell containing a genetic alteration in a *Sal2* gene. For example, the "T-HR mutant specific for a Sal2 mutation" may be the TMD-25 T-HR mutant virus described herein.

"Sal2 mutation", as used herein, refers to a genetic change in the nucleic acid sequence of a Sal2 gene, for example, SEQ ID NO:2 and SEQ ID NO:4, which results in the abnormal proliferation, or predisposition to abnormal proliferation, of a cell carrying such a change. Preferably, the genetic change is a missense mutation. Most preferably, the mutation is a substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

"Uncharacterized abnormally proliferating cell," as used herein, refers to a cell where the cause of the abnormal proliferation is unknown. For example, the genetic alteration that results in abnormal proliferation has not been identified. However, other features of the cell may be characterized.

"Cancer susceptibility gene," as used herein, refers to any gene that, when altered, increases the likelihood that the organism carrying the gene will develop a proliferative disorder during its lifetime. Examples of such genes include proto-oncogenes, tumor suppressor genes, and genes involved in the regulation of cell growth, the cell cycle, and apoptosis.

"Proliferative disease," as used herein, refers to any genetic change within a differentiated cell that results in the abnormal proliferation of a cell. Such changes include mutations in genes involved in the regulation of the cell cycle, of growth control, or of apoptosis and can further include tumor suppressor genes and proto-oncogenes. Specific examples of proliferative diseases are the various types of cancer, such as ovarian cancer. However, proliferative diseases may also be the result of the cell becoming infected with a transforming virus.

"Proliferative disease-associated alteration," as used herein, refers to any genetic change within a differentiated cell that results in the abnormal proliferation of a cell. Preferably, such a genetic change correlates with a

statistically significant (e.g., the p-value is less than or equal to 0.05) increase in the risk of acquiring a proliferative disease. Examples of such genetic changes include mutations in genes involved in the regulation of the cell cycle, of growth control, or of apoptosis and can further include mutations in tumor suppressor genes and proto-oncogenes. A further example of a proliferative disease-associated alteration is the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

"Abnormal proliferation," as used herein, refers to a cell undergoing cell division that normally does not undergo cell division or that undergoes cell division at an increased frequency when compared to a wild-type cell.

The term "alteration," when used herein, in reference to a gene, refers to a change in the nucleic acid sequence. Such a change may include, for example, insertions, deletions, and substitutions of one or more nucleic acids, as well as inversions and duplications.

"Genetic lesion," as used herein, refers to a nucleic acid change. Examples of such a change include single nucleic acid changes as well as deletions and insertions of one or more nucleic acid. However, genetic lesions can also include duplications and inversions. In addition, a genetic lesion may be a naturally-occurring polymorphism, for example, one that predisposes an organism carrying the polymorphism to acquiring a proliferative disease.

"Polymorphism," as used herein, refers to an alteration in a nucleic acid sequence, for example, a gene. Such an alteration may result in a codon change, which in turn may result in, for example, the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

"Modification of function," as used herein, refers to a change in the function of the protein. Such a change can, for example, result in the partial or complete loss of function, but it can also result in a gain of function.

As used herein, the term "promoter" is intended to encompass transcriptional regulatory elements, that is, all of the elements that promote or regulate transcription, including core elements required for basic interactions

between RNA polymerase, transcription factors, upstream elements, enhancers, and response elements.

"Operably linked," as referred to herein, describes the functional relationship between nucleic acid sequences, for example, a promoter sequence, and a gene to be expressed. Operably linked nucleic acids may be part of a contiguous sequence. However a physical link is not necessary for two nucleic acid sequences to be operably linked. For example, enhancers can exert their effect over long distances and therefore do not require a physical link in sequence to the gene whose transcription they affect.

Reference herein to the "transcriptional regulatory elements" of a gene or a class of genes includes both the entire gene as well as an intact region of naturally-occurring transcriptional regulatory elements. Also included are transcription regulatory elements modified by, for example, rearrangement of the elements, deletion of some elements or of extraneous sequences, and insertion of heterologous elements.

The term "knockout," as used herein, refers to an alteration in the sequence of a specific gene that results in a decrease of function of that gene. Preferably the alteration results in undetectable or insignificant expression of the gene and in a complete or partial loss of function. Furthermore, the disruption may be conditional, e.g., dependent on the presence of tetracycline. Knockout animals may be homozygous or heterozygous for the gene of interest. In addition, the term knockout includes conditional knockouts, where the alteration of the target gene can occur, for example, as a result of exposure of the animal to a substance that promotes target gene alteration, introduction of an enzyme that promotes recombination at the target gene site (e.g., Cre in the Cre-lox system, or FLP in the FLP/FRT system), or any other method for directing target gene alteration.

"Conditionally expressed," as used herein, refers to any method that may be used to control expression of a gene, such as a transgene. These methods may, for example, include the use of promoters that are regulated by a substance, such as tetracycline, that can be administered to the organism, or of promoters

that are only active at certain stages of development or in certain tissues. In addition, conditional expression may involve inactivating a gene, for example, by FLP/FRT- or Cre-lox-mediated recombination.

The term "restriction fragment length polymorphism (RFLP) analysis," as used herein, refers to a method of determining whether an organism carries a specific nucleic acid sequence, for example, a specific alteration in a gene. This method may involve, for example, amplification of a nucleic acid from the organism, followed by cleavage of the nucleic acid with an enzyme, such as a restriction enzyme, and visualizing the products of the cleavage reaction. Furthermore, the cleavage products may be compared to control reactions.

As used herein, "alters proliferation" refers to any change in the proliferation of a cell. For example, this term can be used to describe an increase or a decrease in the rate of cell division. In addition, an alteration of proliferation may refer to a normally quiescent cell entering into the cell cycle or a normally dividing cell ceasing to enter into the cell cycle.

"Measuring protein levels," as used herein, includes any standard assay used in the art to either directly or indirectly determine protein levels. Such assays, for example may include the use of an antibody, Western analysis, Bradford assays, and spectrophotometric assays.

"Measuring nucleic acid levels," as used herein, includes any standard assay used in the art to either directly or indirectly determine nucleic acid levels. Such assays include, for example, hybridization analysis, gel electrophoresis, Northern blots, Southern blots, and spectrophotometric assays.

By a "substantially pure polypeptide" is meant a polypeptide (for example, a Sal2 polypeptide) that has been separated from components that naturally accompany it. Typically, the polypeptide is substantially pure when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a Sal2 polypeptide. A substantially pure Sal2 polypeptide may be obtained, for example, by extraction from a natural source (for example, a

mammalian cell); by expression of a recombinant nucleic acid encoding a Sal2 polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

By "isolated DNA" is meant DNA that is free of the genes which, in the naturally-occurring genome of the organism from which the DNA of the invention is derived, flank the gene. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding additional polypeptide sequence.

Advantages

The tumor host range selection procedure described herein has significant advantages over genetic screens and biochemical approaches used in the past to identify viral functions and to elucidate aspects of the interaction between virus and host. For example, previous studies using conditional lethal mutants of the polyoma viruses failed to uncover the large T antigen function involving interaction with mSal2 despite the fact that this interaction is essential for virus growth both *in vitro* (e.g., in tissue culture) and *in vivo* (e.g., in the mouse). In contrast to the directed search for host range mutants based on complementation with integrated viral genes (Benjamin, *Proc. Natl. Acad. Sci. U.S.A.* 67:394-399 (1970)), the 'tumor host range' selection procedure of the invention is an undirected search utilizing non-polyoma transformed or tumor derived cells. Selection of virus mutants is therefore unbiased except for the possibility of being conditional on the transformed state of the particular permissive host being used. Thus, the inventive strategy can lead to the identification of viral functions and cellular targets not revealed by conventional genetic screens or co-immunoprecipitation.

Furthermore, the methods of the invention also have a particular advantage over standard chemotherapy treatments, and the like, in that they are specific for cells with a proliferative disease. Therefore, one would expect this type of therapy to have fewer toxic side effects than the chemotherapeutic agents used today.

Brief Description of the Drawings

Fig. 1 shows photographs of the growth of wild-type polyoma virus and the TMD25 virus on host cells.

Fig. 2A shows the 20 bp sequence duplication responsible for the TMD25 mutation.

Fig. 2B shows the interaction of mSal2 clones with wild-type polyoma virus proteins and the TMD25 virus proteins in a yeast two-hybrid assay.

Fig. 2C shows deletion analysis of the TMD25 mutant.

Fig. 3A shows the regions of the *mSal2* gene used to develop antibodies.

Fig. 3B shows antibody detection of $p150^{sal2}$ as the *mSal2* gene product in a Western blot using protein from mouse and human cells.

Fig. 3C shows a Western blot of extracts from human 293 and U2OS cells that was first probed with an antiserum against the mSal2 carboxyl-terminus. The filter was then stripped and re-probed with an antibody against the mSal2 amino-terminus.

Fig. 4A shows the binding of mSal2 to wild-type polyoma virus but not to TMD25 large T protein *in vitro*.

Fig. 4B shows the binding of mSal2 and wild-type, but not TMD-25 mutant, large T protein in transfected 3T3 cells. These results are confirmed in BMK cells infected with wild-type polyoma virus and with TMD25 mutant virus.

Fig. 5A shows the failure of TMD25 to replicate in newborn mice.

Fig. 5B shows that TMD25 fails to replicate in BMK cells and that $p150^{sal2}$ represses viral origin replication.

Fig. 6 shows a Western blot of mSal2 expression in various mouse tissues.

Fig. 7 shows a Western blot of hSal2 expression in human ovarian tumors.

Fig. 8 shows expression of $p150^{sal2}$ in human 293 cells.

Fig. 9 shows immunostaining of $p150^{sal2}$ in human ovary tissue (A) and in ovarian tumors (B).

Fig. 10A shows that $p150^{sal2}$ suppresses growth of human ovarian tumor cells, which is indicated by a reduction in BrdU incorporation in $p150^{sal2}$ transfected cells.

Fig. 10B shows a colony reduction assay that indicates that cells transfected with $p150^{sal2}$ are less viable than control transfected cells.

Fig. 11 is an agarose gel showing that the 73S allele is lost in some ovarian tumors.

Detailed Description of the Invention

The present invention provides a method for identifying genes that play a role in cancer as well as methods for diagnosing and treating patients who have cancers involving these genes.

Identifying genes altered in cancerous cells

Host range selection of viruses

The present invention describes the use of tumor host range mutant viruses (T-HR mutants) that are capable of replicating in abnormally proliferating cells but not in normal cells. Therefore, these viruses are useful for identifying genes altered in abnormally proliferating cells. T-HR mutants generally have a mutation that causes a modification of function of the protein encoded by that gene. These mutations typically lie in the transforming genes of the DNA tumor viruses and are usually activators of cellular proto-oncogenes or inactivators of tumor suppressor genes. T-HR mutants may be isolated based on their ability to propagate (i.e. to replicate and disseminate) only in tumor cells that have a mutation in a cellular protein that is normally targeted by a viral transforming protein.

The methods of the invention have been applied to a 'tumor host range' selection procedure using the polyoma virus as a tool to search for new interactions of viral proteins, e.g., T antigens, with cellular proteins. The rationale behind this approach is based on the idea that genetic changes in tumor cells resulting in a modification of function of the cellular protein can provide the basis for a search to uncover new viral functions and interactions with cellular targets. In principle, 'Tumor host range' selection could reveal mutations in other functions, e.g., VP1, 2 or 3 involving interactions with receptors or the cellular machinery involved in virus uptake, uncoating or transport to the nucleus, or even in some aspect of virus assembly, or enhancer mutations that lead to alterations in enhancer function.

For example, alterations in yet unknown targets of viral genes might occur in spontaneous tumors or non-virally transformed cells. This suggests a rationale for isolating T-HR mutants based on modification of function in cancer cells. Mutants selected to grow in tumor cells, but not in normal cells, are useful for identifying new viral gene functions and their cellular targets. Targets identified in this way may include products of tumor suppressor genes or proto-oncogenes or any factor expressed in normal cells, which the virus must inactivate in order to propagate, but that is no longer expressed in tumor cells.

Identification of mSal2

The utility of the T-HR mutant based approach for identifying new genes involved in the susceptibility to proliferative diseases is shown by the identification of *mSal2*. The use of a T-HR mutant coupled with the power of the yeast two-hybrid screen resulted in the identification of a cellular target protein. Using T-HR mutants to identify cell cycle regulatory proteins is advantageous on two levels; first, in choosing an appropriate wild-type 'bait' corresponding to the region altered in the mutant, and second, in enabling a counterscreen where lack of interaction with the mutant is helpful in identifying cellular target(s) relevant to the mutant phenotype and possibly also to the transformed state of the

permissive host. One embodiment of the general protocol included as an aspect of the invention is outlined in Table 1 below.

Table 1. Tumor Host Range Mutants - Selection Procedure and Target Identification

I. Mutant Selection

1. Random mutagenesis of wild-type viral DNA
2. Amplification of the mutant virus by growth in tumor cells
3. Cloning by plaque isolation on tumor cells
4. Screening of plaque lysates for the absence of growth in normal cells
5. Molecular cloning and sequencing of the mutant viral DNA

II. Target Identification and Validation

6. Screening of a mouse embryo cDNA library in yeast with wild-type bait
7. Counterscreening positive clones for lack of interaction with mutant bait
8. Construction of complete cDNA expressing the target protein
9. Verification of viral protein-cellular target interactions *in vitro* and *in vivo* (e.g., T antigen-cellular protein interactions).

III. Identification of Risk Factors

10. Sequencing DNA derived from a tumor
11. Sequencing DNA derived from normal tissue of the same patient
12. Using the sequence information to establish whether the mutation is somatic or germline
13. Using this information in an epidemiological study to assess risk factors in a population

What follows is an illustration of the use of the methods of the invention to identify a new target of large T antigen, referred to as mSal2, using T-HR mutants of the polyomavirus. First, tumor host range selection identified a host range mutant of the polyomavirus that is able to grow in certain tumor or

transformed cells but not in normal cells. The mutant virus encodes an altered large T antigen protein and is defective in replication and tumor induction in newborn mice. Next, mSal2 was identified as a binding target of the polyoma virus large T antigen through a yeast two-hybrid screen. mSal2 shows no interaction with the mutant large T antigen. Specifically, the mutant virus fails to bind mSal2 and is unable to propagate or to induce most of the tumor types in the mouse that the wild-type virus typically induces.

The gene product $p150^{sal2}$ is expressed in a number of mouse and human tissues. It is found in nuclei of germinal epithelial cells from normal human ovary but is missing or altered in ovarian carcinomas derived from these cells (Table 3). Using an antibody to mSal2 that cross-reacts with the human protein, Sal2 was shown to be expressed as a protein of approximately 150 kDa in several normal murine and human tissues. Normal human ovarian epithelial cells show strong nuclear staining with the antibody. A majority of ovarian carcinomas derived from these cells show no detectible $p150^{sal2}$ by Western analysis and are negative by *in situ* immunochemistry. Some tumors display diffuse cytoplasmic, rather than nuclear, staining. (See Examples below.)

mSal2 is a zinc finger protein and a putative transcription factor that may have a role as a tumor suppressor. *mSal2* is homologous to the *Drosophila* homeotic gene *spalt* and to *sal* homologues identified in several vertebrate species (see below). The human homologue of the *Drosophila spalt* gene, *hSal2*, has been mapped adjacent to, or overlapping with, a chromosomal region associated with a loss of homozygosity in ovarian and other cancers.

The *spalt* or *sal* gene family of transcription factors is conserved in evolution from flies to man. First identified in *Drosophila*, *spalt* is a region-specific homeotic gene which functions in specifying anterior and posterior structures in the early embryo (Kuhnlein et al., *EMBO J* 13:168-179 (1994); Jurgens et al., *EMBO J* 7:189-196 (1988)) and also in later stages of organogenesis (Kuhnlein et al., *Mech. Dev.* 66:107-118 (1997); Barrio et al., *Dev. Biol.* 215:33-47 (1999)). *spalt*-related *sal* genes have been identified and studied in worms (Basson et al., *Genes Dev.* 10:1953-1965 (1996)), fish (Koster et al.,

Development 124:3147-3156 (1997)), frogs (Hollemann et al., *Mech. Dev.* 55:19-32 (1996); Onuma, *Biochem. Biophys. Res. Commun.* 264:151-156 (1999)), mice (Ott et al., *Mech. Dev.* 56:117-128 (1996); Kohlhase et al., *Nat. Genet.* 18:81-83 (2000)) and man (Kohlhase et al., *Genomics* 38:291-298 (1996); Kohlhase et al., *Genomics* 1:216-222 (1999); Kohlhase et al., *Cytogenet. Cell Genet.* 84:31-34 (1999)). In humans, a defect in the *hSall* gene underlies the multiple developmental defects seen in Townes-Brocke syndrome (Kohlhase et al., *Nat. Genet.* 18:81-83 (1998)). Sal proteins contain multiple Zinc fingers, which frequently occur as C2H2 pairs with a conserved motif (Kuhnlein et al., *EMBO J* 13:168-179 (1994)). mSal2 has a structural arrangement typically seen in vertebrates with a single finger (C3H) near the amino-terminus and a cluster of three fingers (C2H2) considered essential for DNA binding in the middle portion of the protein (Pabo et al., *Annu. Rev. Biochem.* 61:1053-1095 (1992)). Like other Sal proteins, mSal2 has both glutamine-rich and proline- and alanine-rich sequences consistent with its transcriptional activator and repressor functions.

Although it has been shown in several species that Sal family transcription factors play important roles in embryonic development, downstream target genes have yet to be identified. Nevertheless, two important signaling pathways lying upstream of *sal* have been recognized. Regulation of *spalt* occurs in part through *dpp*, a member of the TGF-β family, which functions as a 'gradient morphogen' in the early *Drosophila* embryo (de Celis et al., *Nature* 381:421-424 (1996); Lecuit et al., *Nature* 381:387-393 (1996); Nellen et al., *Cell* 85:357-368 (1996)). In *Medaka*, *Sall* expression occurs in response to *hh* (*hedgehog*) and is downregulated through PK-A (Koster et al., *Development* 124:3147-3156 (1997)). The TGF-β family of polypeptides has well known inhibitory effects on epithelial cell growth and survival. Disruptions in signaling pathways initiated by TGF-β are known to occur in some cancers (Kretzschmar et al., *Current Opinion in Genetics & Development* 8:103-111 (1998); Serra et al., *Nature Med.* 2:390-391 (1996)). In particular, mutations in *SMAD* genes, essential mediators of signaling via TGF-β receptors, have been linked to pancreatic, colorectal, and other cancers (Eppert et al., *Cell* 86:543-552 (1996);

Hahn et al., *Science* 271:350-353 (1996); Schutte et al., *Cancer Res.* 56:2527-2530 (1996)). Similarly, disruptions in signaling *via* '*hedgehog*' ligands and their '*patched*' receptors are important in development of basal cell carcinoma (Hahn et al., *Cell* 85:841-851 (1996); Johnson et al., *Science* 272:1668-1671 (1996); Oro et al., *Science* 276:817-821 (1997); Stone et al., *Nature* 384:129-134 (1996)).

Diagnosis

Diagnosis and Risk Assessment

In addition to helping identify genes that are altered in cancerous cells, target gene profiles can also be used to diagnose and/or stage various proliferative disorders and for diagnosing pre-symptomatic genetic lesions in normal tissues. The methods of the present invention can be used to diagnose cancerous cells in a patient by determining whether the cells of the patient can act as permissive hosts for the growth of a mutant virus, particularly a T-HR mutant. As described above, a permissive host for the growth of a mutant virus (e.g., a mutant virus that lacks a functioning transforming protein) has a mutation in a cellular gene that is the target for the wild-type viral protein corresponding to the mutant viral protein. This cellular mutation is believed to compensate for the modification of function in a particular gene in the T-HR mutant and contribute to the cancerous phenotype of the cell.

Once a target protein has been identified, tests for the lack of interaction of the cellular protein with the mutant viral protein are used to confirm the specificity of the interaction of the cellular protein with the wild-type (transforming) protein. A lack of interaction indicates that binding of the wild-type viral protein to the cellular protein is specific. Protein interaction can be verified by numerous methods know to those skilled in the art, including, for example, yeast two-hybrid assays, GST-pull down assays, co-immunoprecipitation, and Far-Western analysis. General guidance regarding these techniques can be found in standard laboratory manuals, such as Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York,

NY, (1994)), and Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, N.Y., (1989)). Once an interaction between the wild-type viral protein and the cellular protein is confirmed, the complete gene and gene product can readily be identified by those skilled in the art using, for example, the methods described below.

The present invention recognizes that the T-HR mutant selection procedures identified herein may identify mutant cellular genes, and their encoded protein products, e.g., cellular genes encoding cell cycle proteins, tumor suppressors, proto-oncogenes, transcriptional factors, regulators of apoptosis, etc., that have genetic lesions associated with a particular proliferative disorder. Those skilled in the art will appreciate that many proliferative disorders, such as cancers, correlate with a particular mutation or mutations in the DNA of a patient. By comparing the sequence for a particular gene in both normal and tumor tissue from the same patient, one can determine if the mutation is of somatic or germline origin. This information that may be used to screen a population as a whole for individuals that are at an increased risk of developing a particular type of proliferative disorder.

The present invention provides a method of identifying a genetic lesion in a cell by determining whether a cell can act as a permissive host for the growth of a particular T-HR mutant, such a T-HR mutant virus being capable of growing on a cell having a specific genetic lesion and not being capable of growth on a cell lacking this genetic lesion. This type of information may even be used to further characterize the cancer cell (e.g., to grade the stage to which the cancer has progressed).

In addition, the cellular gene that encodes a protein that is a target for a viral transforming protein may also be analyzed to determine whether there is a genetic lesion in the cellular gene. Such a genetic lesion may be associated with a particular cancer. As noted above, the present inventors describe a genetic lesion that may be associated with ovarian cancer has been identified in a *Sal2* gene. Specifically, this genetic lesion, corresponding to the substitution of a Cys for the Ser at position 73 in protein encoded by the *mSal2* gene of SEQ ID NO:4,

has been identified in DNA from blood samples from patients with ovarian cancer. Probes and primers based on this genetic lesion may be used as markers to detect the Ser73Cys change in samples from other patients.

A genetic lesion in a candidate gene may be identified in a biological sample obtained from a patient using a variety of methods available to those skilled in the art. Generally, these techniques involve PCR amplification of nucleic acid from the patient sample, followed by identification of the genetic lesion by either altered hybridization, aberrant electrophoretic gel migration, restriction fragment length polymorphism (RFLP) analysis, binding or cleavage mediated by mismatch binding proteins, or direct nucleic acid sequencing. Any of these techniques may be used to facilitate detection of a genetic lesion in a candidate gene, and each is well known in the art; examples of particular techniques are described, without limitation, in Orita et al. (*Proc. Natl. Acad. Sci. USA* 86:2766-2770, 1989) and Sheffield et al. (*Proc. Natl. Acad. Sci. USA* 86:232-236 (1989)). Furthermore, expression of the candidate gene in a biological sample (e.g., a biopsy) may be monitored by standard Northern blot analysis or may be aided by PCR (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994); *PCR Technology: Principles and Applications for DNA Amplification*, H.A. Ehrlich, Ed., Stockton Press, NY; Yap et al., *Nucl. Acids. Res.* 19:4294 (1991)).

Once a genetic lesion is identified using the methods of the invention (as is described above), the genetic lesion is analyzed for association with an increased risk of developing a proliferative disorder. In this respect, the present invention provides a method of detecting the presence of a genetic lesion in a human *Sal2* gene in a physiological sample, however the method is not limited to the *Sal2* gene, but rather can be applied to any gene that is associated with an increased risk for developing a proliferative disorder.

Furthermore, antibodies against a protein produced by the gene included in the genetic lesion, for example the Sal2 protein. Antibodies may be used to detect altered expression levels of the protein, including a lack of expression, or a change in its mobility on a gel, indicating a change in structure or size. In

addition, antibodies may be used for detecting an alteration in the expression pattern or the sub-cellular localization of the protein. Such antibodies include ones that recognize both the wild-type and mutant protein, as well as ones that are specific for either the wild-type or an altered form of the protein, for example, one encoded by a polymorphic *Sal2* gene. Monoclonal antibodies may be prepared using the Sal2 proteins described above and standard hybridoma technology (see, e.g., Kohler et al., *Nature* 256:495 (1975); Kohler et al., *Eur. J. Immunol.* 6:511 (1976); Kohler et al., *Eur. J. Immunol.* 6:292 (1976); Hammerling et al., In *Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, New York, NY (1981); Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)). Once produced, monoclonal antibodies are also tested for specific Sal2 protein recognition by Western blot or immunoprecipitation analysis (by the methods described in, for example, Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)).

Antibodies used in the methods of the invention may be produced using amino acid sequences that do not reside within highly conserved regions, and that appear likely to be antigenic, as analyzed by criteria such as those provided by the Peptide Structure Program (Genetics Computer Group Sequence Analysis Package, Program Manual for the GCG Package, Version 7, 1991) using the algorithm of Jameson and Wolf (*CABIOS* 4:181 (1988)). These fragments can be generated by standard techniques, e.g., by the PCR, and cloned into the pGEX expression vector (Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)). GST fusion proteins are expressed in *E. coli* and purified using a glutathione agarose affinity matrix as described in Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY, (1994)).

To generate rabbit polyclonal antibodies, and to minimize the potential for obtaining antisera that is non-specific, or exhibits low-affinity binding, two or three fusions are generated for each protein, and each fusion is injected into at least two rabbits. Antisera are raised by injections in series, preferably including

at least three booster injections. These methods for antibody production and characterization are applicable to any other protein that is identified by the methods of the invention.

The antibody may be used in immunoassays to detect or monitor protein expression, e.g., Sal2 protein expression, in a biological sample. A polyclonal or monoclonal antibody (produced as described above) may be used in any standard immunoassay format (e.g., ELISA, Western blot, or RIA) to measure polypeptide levels. These levels may be compared to normal levels. Examples of immunoassays are described, e.g., in Ausubel et al. (*Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)). Immunohistochemical techniques may also be utilized for protein detection. For example, a tissue sample may be obtained from a patient, sectioned, and stained for the presence of Sal2 using an anti-Sal2 antibody and any standard detection system (e.g., one which includes a secondary antibody conjugated to horseradish peroxidase). General guidance regarding such techniques can be found in, e.g., Bancroft and Stevens (*Theory and Practice of Histological Techniques*, Churchill Livingstone (1982); and Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, NY (1994)).

Use of hSal2 as a Diagnostic Tool

As an example of the utility of this approach, the likelihood that *hSal2* functions as a tumor suppressor for ovarian cancer has been explored directly by screening a number of ovarian carcinomas for expression of $p150^{sal2}$ and for mutations in the gene. Approximately 80% of the tumors examined were negative or showed altered or reduced patterns of expression by Western analysis. Immunolocalization in frozen tissue sections showed strong staining in nuclei of epithelial cells on the surface of the normal ovary. In most instances, tumor cells showed a complete lack of staining. However, when staining was present in otherwise negative tumors, cytoplasmic rather than nuclear staining was seen in some areas.

Further evidence for *hSal2* function as a tumor suppressor comes from a limited screen for mutations in *hSal2*, which uncovered point mutations in four cases. In addition, cytogenetic approaches and sequencing efforts utilizing microsatellite markers have been used to map *hSal2* adjacent to, and possibly overlapping with, a chromosomal region associated with loss of homozygosity in ovarian (Bandera et al., *Cancer Res.* 57:513-515 (1997)) and other cancers, e.g., bladder cancer (Chang et al., *Cancer Res.* 55:3246-3249 (1995)). Such approaches may continue to be used to map *hSal2* more precisely.

The *mSal2* gene identified by the present invention may further be used to elucidate the cellular pathways of tumor suppression that regulate key cell cycle events. Alternatively, *mSal2* may be used to screen for potential tumors, e.g., lung tumors, brain tumors, stomach tumors, prostate tumors; ovarian tumors, tumors in SCID mice, as well as in knockout or transgenic animals, as discussed in detail below.

Treatment

In addition to providing a method for identifying genes altered in cancer cells and diagnosing patients who carry such mutation, the invention further provides a method of killing an abnormally proliferating cell using a tumor host range mutant virus.

For example, T-HR mutants may be used to specifically target and kill cancer cells in an organism. Since these viruses can only propagate in cells that carry a mutation in a cellular gene that the virus would normally have to activate, in the case of proto-oncogene, or inactivate, in the case of a tumor suppressor gene, in order to propagate, such a virus would be specific to abnormal cells. Therefore, T-HR mutants can be used to specifically eliminate cancer cells from a patient. For example, a T-HR mutant (i.e., a polyomavirus carrying an altered large T antigen causing it to be defective in replication and tumor induction) may be used to selectively kill human ovarian cancer cells that carry a genetic lesion in the *hSal2* gene, such as the Ser73Cys substitution described above.

However, one skilled in the art would realize that any number of genes, including ones involved in cell growth, cell cycle regulation, and apoptosis, may be altered in cancer cells. The methods of the invention are applicable to any alteration in a cancer cell that allows a T-HR mutant to grow. Therefore, any cancer that enables a T-HR mutant to propagate can be treated according to the methods of the invention disclosed herein.

The therapeutic T-HR mutant may be administered by any of a variety of routes known to those skilled in the art, such as, for example, intraperitoneal, subcutaneous, parenteral, intravenous, intramuscular, or subdermal injection. However, the T-HR mutant may also be administered as an aerosol, as well as orally, nasally, or topically. Standard concentrations used to administer a T-HR mutant include, for example, 10^2, 10^3, 10^4, 10^5, or 10^6 plaque forming units (pfu)/animal, in a pharmacologically acceptable carrier. Appropriate carriers or diluents, as well as what is essential for the preparation of a pharmaceutical composition are described, e.g., in *Remington's Pharmaceutical Sciences* (18^{th} edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA, a standard reference book in this field.

Formulations for parenteral administration may, for example, contain excipients, sterile water, or saline. For inhalation, formulations may contain excipients, such as lactose. Furthermore, aqueous solutions may be used, for example, for administration in the form of nasal drops, or as a gel for topical administration. The exact dosage used will depend on the severity of the condition (e.g., the size of the tumor), or the general health of the patient and the route of administration. The T-HR mutant may be administered once, or it may be repeatedly administered as part of a regular treatment regimen over a period of time.

Compounds that may be tested for an effect on proliferative diseases can be from natural as well as synthetic sources. Those skilled in the field or drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the methods of the invention. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-,

prokaryotic-, or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceanographics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries may be produced, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

Transgenic and Knockout Animals

The present invention provides transgenic and knockout animals that develop ovarian tumors and accurately recapitulate many of the features of the human ovarian tumor, an important contribution, since animal models of ovarian carcinoma are currently not available. Without limitation, particularly preferred transgenic or knockout animals are those in which the tumorigenic phenotype is fully penetrant, the rate of progression of the neoplasm is rapid, and/or the lifespan of the transgenic or knock-out animal is not shortened by a knockout- or transgene-related pathology in other organs. Of course, it will be appreciated that these traits are not required.

The generation of transgenic or knockout mice may provide a valuable tool for the investigation of human ovarian cancer by generating a mouse model for studying the disease, based on the description of the human *Sal2* gene provided above. Preferably, the *hSal2* gene is used to produce the transgenic mice or the *mSal2* gene is the target of the knockout. However, other *Sal2* genes

may also be used to produce transgenic mice provided that they are compatible with the mouse genome and that the protein encoded by this gene is able to carry out the function of the mSal2 protein.

Furthermore, a transgene, such as a mutant *Sal2* gene, may be conditionally expressed (e.g., in a tetracycline sensitive manner). For example, the promoter for the *Sal2* gene may contain a sequence that is regulated by tetracycline and expression of the Sal2 gene product ceases when tetracycline is administered to the mouse. In this example, a tetracycline-binding operator, tetO, is regulated by the addition of tetracycline, or an analog thereof, to the organism's water or diet. The tetO may be operably-linked to a coding region, for example a mutant *Sal2* gene. The system also may include a tetracycline transactivator (tTA), which contains a DNA binding domain that is capable of binding the tetO as well as a polypeptide capable of repressing transcription from the tetO (e.g., the tetracycline repressor (tetR)), and may be further coupled to a transcriptional activation domain (e.g., VP16). When the tTA binds to the tetO sequences, in the absence of tetracycline, transcription of the target gene is activated. However, binding of tetracycline to the tTA prevents activation. Thus, a gene operably-linked to a tetO is expressed in the absence of tetracycline and is repressed in its presence. The tetracycline regulatable system is well known to those skilled in the art and is described in, for example, WO 94/29442, WO 96/40892, WO 96/01313, and Yamamoto et al. (*Cell* 101:57-66 (2000).

In addition, the knockout organism may be a conditional knockout. For example, FRT sequences may be introduced into the organism so that they flank the gene of interest. Transient or continuous expression of the FLP protein may then be used to induce site-directed recombination, resulting in the excision of the gene of interest. The use of the FLP/FRT system is well established in the art and is described in, for example, U.S. Patent Number 5,527,695, and in Lyznik et al. (*Nucleic Acid Research* 24:3784-3789 (1996)).

Conditional knockout organisms may also be produced using the Cre-lox recombination system. Cre is an enzyme that excises DNA between two recognition sites termed loxP. The *cre* transgene may be under the control of an

inducible, developmentally regulated, tissue specific, or cell-type specific promoter. In the presence of Cre, the gene, for example a *Sal2* gene, flanked by loxP sites is excised, generating a knockout. This system is described, for example, in Kilby et al. (Trends in Genetics 9:413-421 (1993)).

Particularly preferred is a mouse model for ovarian cancer wherein the nucleic acid encoding a *Sal2* gene is expressed in the cells of the ovary of the transgenic mouse such that the transgenic mouse develops ovarian tumors. The mice preferably contain a large T antigen transgene, such as one expressing an appropriate (carboxyl-terminal) fragment of large T antigen under the control of an ovarian specific promoter, or have a knockout of the *mSal2* gene. In addition, ovarian cell lines from these mice may be established by methods standard in the art.

Transgenic animals may be made using standard techniques. For example, a gene encoding a cellular proto-oncogene, tumor suppressor gene, or other cellular protein, e.g., a cell cycle regulating protein, may be provided using endogenous control sequences or using constitutive, tissue-specific, or inducible regulatory sequences. Any tissue specific promoter may direct the expression of any Sal2 protein used in the invention, such as ovarian specific promoters, bladder specific promoters, and colon specific promoters. For example, knockout mutations may be engineered in the gene encoding the proto-oncogene or tumor suppressor gene and the mutated gene may be used to replace the wild-type *Sal2* gene.

Construction of transgenes can be accomplished using any suitable genetic engineering technique, such as those described in Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, N.Y., (1989)). Many techniques of transgene construction and of expression constructs for transfection or transformation in general are known and may be used for the disclosed constructs. Although the use of *hSal2* in the transgene constructs is used as an example, any other protein encoded by an oncogene may also be used.

One skilled in the art will appreciate that a promoter is chosen that directs expression of the oncogene in the tissue in which cancer is expected to develop. For example, as noted above, any promoter that regulates expression of *hSal2* in ovarian cancer cells can be used in the expression constructs of the present invention. Preferred ovarian promoters include, for example, promoters that are expressed in ovarian epithelial cells, such as, the polyoma virus promoter, the SPARK promoter, and the DOC-2 promoter. One skilled in the art would be aware that the modular nature of transcriptional regulatory elements and the absence of position-dependence of the function of some regulatory elements, such as enhancers, make modifications such as, for example, rearrangements, deletions of some elements or extraneous sequences, and insertion of heterologous (i.e., foreign) elements possible. Numerous techniques are available for dissecting the regulatory elements of genes to determine their location and function. Such information can be used to direct modification of the elements, if desired. It is preferred, however, that an intact region of the transcriptional regulatory elements of a gene is used. Once a suitable transgene construct has been made, any suitable technique for introducing this construct into embryonic cells can be used, an example of such a technique is provided in Example 9.

Animals suitable for transgenic experiments can be obtained from standard commercial sources such as Taconic (Germantown, NY). Many strains are suitable, but Swiss Webster (Taconic) female mice are preferred for embryo retrieval and transfer. B6D2F (Taconic) males can be used for mating and vasectomized Swiss Webster studs can be used to stimulate pseudopregnancy. In addition, vasectomized mice and rats are also publicly available from the above-mentioned suppliers. However, one skilled in the art would also know how to make a transgenic mouse or rat. An example of a protocol that can be used to produce a transgenic animal is provided in Example 9.

Use of Transgenic and Knockout Animals

The disclosed transgenic and knockout animals may be used as research tools to determine genetic and physiological features of a cancer, and for identifying compounds that can affect ovarian and other tumors. Knockout animals also include animals where the normal gene has been inactivated or removed and replaced with a mutant form of this gene, for example, a polymorphic allele. These animals can serve as a model system for assessing the risk of acquiring a proliferative disease that is associated with a particular mutation.

In general, the method of identifying markers associated with a proliferative disorder, such as ovarian tumors, involves comparing the presence, absence, or level of expression of genes, either at the RNA level or at the protein level, in tissue from a transgenic or knockout animal as described above, and in tissue from a matching non-transgenic or knockout animal. Standard techniques for detecting RNA expression, e.g., by Northern blotting, or protein expression, e.g., by Western blotting, are well known in the art. Differences between animals such as the presence, absence, or level of expression of a gene indicate that the expression of the gene is a marker associated with a proliferative disorder, such as ovarian tumors. The molecular markers, once identified, can be used to predict whether patients with carcinoma will have indolent or aggressive disease, and may be mediators of disease progression. Identification of such mediators would be useful since they are possible therapeutic targets. Identification of markers can take several forms.

One method by which molecular markers may be identified is by use of directed screens. Patterns of accumulation of a variety of molecules that may regulate growth can be surveyed using immunohistochemical methods. Screens directed at analyzing expression of specific genes or groups of molecules implicated in pathogenesis can be continued during the life of the transgenic or knockout animal. Expression can be monitored by immunohistochemistry as well as by protein and RNA blotting techniques. Mestastatic foci, once formed, can also be subjected to such comparative surveys.

Alternatively, molecular markers may be identified using genomic screens. For example, ovarian tissue can be recovered from young transgenic or knockout animals (e.g, that may have early stage carcinoma) and older transgenic or knockout animals (e.g., that may have advanced stage carcinoma), and compared with similar material recovered from age-matched normal littermate controls to catalog genes that are induced or repressed as disease is initiated, and as disease progresses to its final stages. These surveys will generally include cellular populations in the ovary.

This analysis can also be extended to include an assessment of the effects of various treatment paradigms (including the use of compounds identified as affecting ovarian tumors in the transgenic or knockout animals) on differential gene expression (DGE). The information derived from the surveys of DGE can ultimately be correlated with disease initiation and progression in the transgenic or knockout animals.

The following examples are meant to illustrate the invention and should not be construed as limiting.

Examples

Example 1: Isolation Of TMD-25 Using A 'Tumor Host Range' Selection

A procedure for isolating 'tumor host range' mutants (e.g., T-HR mutants) and identifying cellular targets is outlined below.

Identification of a Host Factor that interacts with T Antigens

1) Select Host Range Mutants
2) Identify Host Range Mutations
3) Identify Host Range Target and Validation
4) Biological Properties:
    (i) Viral DNA Replication
    (ii) Transformation
    (iii) Tumorigenicity

Permissive hosts were chosen based on a screen of mouse cell lines derived from non-polyoma-induced tumors or transformed cells using the following criteria: (i) susceptibility to lytic infection by wild-type polyoma virus, and (ii) ability to be used in standard plaque assays.

Among a number of qualifying cell lines, two were chosen: A6241, derived from a spontaneous mammary tumor in a C57BR mouse, and TCMK-1, a SV40-transformed baby mouse kidney cell line. Primary baby mouse kidney epithelial cells (BMK) were used throughout as the non-permissive host.

Randomly mutagenized virus was prepared by passage of a plasmid containing wild-type polyoma viral DNA through the error prone Mut D strain of *E. coli*, followed by excision of the viral genome and transfection into TCMK-1 cells. After several cycles of virus growth in the same cells, individual plaques were isolated using TCMK-1 cells. An aliquot of virus in each plaque suspension was inoculated into BMK cell cultures. Virus from plaques that induced no cytopathic effect (CPE) on BMK cells after 10-14 days was amplified using TCMK-1 cells. Mutant DNAs were cloned, reconstituted as virus by transfection of permissive cells, and confirmed to retain the desired host range. The frequency of mutants was approximately one in several thousand plaques tested. The T-HR mutant TMD-25 was isolated by this procedure.

Fig.1 shows the results of CPE tests comparing wild-type polyoma virus and TMD-25 growth in BMK, TCMK-1, and A6241 cells. Primary baby mouse kidney cells (BMK), SV40 Large T antigen transformed mouse kidney cells (TCMK), and spontaneous mouse mammary tumor cells (A6241) were mock-infected (Mock), or infected with 2-5 pfu of wild-type polyoma virus (PTA) or of T-HR mutant TMD25. The photographs were taken 14 days post infection and show the different cytopathic effects of viral growth.

TMD25 mutants grew poorly, if at all, on primary BMK cells, but could grow on transformed or tumor-derived cells, while wild-type polyoma virus grew well on all three cell-types. Extensive CPE developed in the TCMK-1 and A6241 cultures infected by the TMD25 mutant. Infectious mutant virus was produced in these cultures, although with somewhat slower kinetics and with

lower final yields compared to wild-type virus. In contrast, no discernible CPE was noted in mutant-infected BMK cultures, even after extended periods of incubation of up to three weeks. Growth of TMD-25 on the spontaneous tumor line A6241 rules out the possibility that its growth depends strictly on complementation by SV40 large T antigen, which is expressed in TCMK-1.

Example 2: Sequencing Of TMD-25 And Screening For Targets In Yeast

The mutation in TMD-25 responsible for its 'tumor host range' was localized to the carboxyl-terminal half of polyoma large T antigen as a result of studies using chimeric viruses constructed by ligating complementary DNA fragments from TMD-25 and wild-type virus. A combination of marker rescue and sequence analysis of this region revealed a twenty base pair duplication (circled) in TMD-25 encompassing the carboxyl-terminus of large T antigen. The resulting frameshift leads to replacement of the last 12 amino acids by 11 foreign residues (underlined) (SEQ ID NOS:9 to 12) (Fig. 2A).

It is possible that the carboxyl-terminal region of large T antigen is involved in binding to some cellular target as an essential step in virus growth and that the mutation in TMD-25 abolishes this interaction. As a first step toward identifying a possible cellular target, a cDNA library constructed from 9.5 to 10.5 day-old mouse embryos was screened in yeast two-hybrid assays, using the carboxyl-terminal portion of normal large T antigen (amino acids 335-782) as bait.

Twenty-two positive clones were analyzed. Nineteen of these clones were represented by nine independent but overlapping cDNA sequences that centered around a sixty-six amino acid region (amino acids 900-965) encompassing a zinc finger pair in the carboxyl-terminal region of the mSal2 protein cDNAs (Fig. 2B, Left Panel, and discussed below). The identified sequences showed strong homology to the human gene *hSal2*, which is related to *spalt* in *Drosophila*.

The positive mSal2 clones did not interact with the carboxyl-terminus of TMD25 large T antigen, as indicated by the growth (+) of yeast colonies on

histidine minus plates when using normal polyoma large T antigen as bait, but no growth using TMD25 large T antigen as bait (Fig. 2B, Right Panel), consistent with the notion that the host range defect of TMD-25 is based on its inability to bind this protein. All the His+ yeast colonies were also LacZ positive.

On continuous propagation in permissive cells, the TMD-25 mutant proved to be unstable, giving rise to wild-type virus revertants. To obtain a stable mutant and to further pinpoint the region of large T antigen essential for binding, (SEQ ID NOS:13 to 21), an analysis of the wild-type bait construct was carried out using mSal2 interaction in yeast as an assay (Fig. 2C). Truncation of the last six amino acids had no perceptible effect, but further truncations into the P-L-K sequence at positions 774-776 resulted in a loss of interaction. A deletion of these three amino acids in the context of an otherwise intact large T antigen was sufficient to prevent interaction with mSal2 and to recreate the host range phenotype shown in Fig. 1. The large T antigen deletion mutant 774-776 is hereafter referred to as TMD-25. The original defect of TMD25 is underlined, and the three amino acid region is framed in Fig. 2.

Example 3: Validation Of mSal2 As A Target Of Large T Antigen

A complete cDNA was obtained using RACE. The sequence was found to be identical to that reported recently for mSal2, with a Glu rather than a Lys residue at position 350. The genomic sequence indicates two alternate short 5' exons each encoding 24 amino acids and one unique 3' exon encoding 980 amino acids. The overall homology with hSal2 is 85% using the Blast 2 Sequence program. Eight Zinc fingers are apparent in exon 2. These zinc fingers are organized in four groups with the carboxyl-terminal pair presumed to be an essential part of the large T antigen interaction domain (Figs. 2B and 3A). Fig. 3A shows the corresponding gene region of the mSal2 protein fragments used to develop antibodies. The exons are boxed, with the zinc fingers represented as stripes.

Fig. 3B shows the antibody detection of *in* vitro translated full-length mSal2 and $p150^{sal2}$ in mouse and human cells. A polyclonal antibody was made

in rabbits against a GST fusion protein containing 131 amino acids from the carboxyl-terminal large T antigen interaction domain. Extracts of mouse 624 and human 293 cell lines probed with this antibody show a single protein species migrating at approximately 150 kDa (Fig. 3B, Right Panel). A monoclonal antibody against a 108 amino acid amino terminal fragment spanning exons 1 and 2 was isolated (Fig. 3A). This antibody also detected mSal2 as a 150 kDa *in vitro* translation product (Tr), as well as a protein present in normal mouse brain extracts (Br)(Fig. 3B). This gene product of mouse and human origin is referred to as $p150^{sal2}$. To confirm that the single band from the human cell extract is hSal2, extracts from two human cell lines first were probed with the polyclonal antibody made against the carboxyl-terminus of mSal2. The filter was then stripped and reprobed with the anti-mSal2 amino-terminus polyclonal antibody. The identical band was detected with each of the two antibodies in the human cell lysates (Fig. 3C).

In vitro pull down assays were carried out using a GST fusion of the large T antigen interaction domain of $p150^{sal2}$ and extracts of lytically infected or transfected cells (Fig. 4A). The filter was blotted with an anti-large T antigen antibody. Lanes "a" to "c" show pulldown assays using wild-type polyoma, lytic infected BMK cells: lane "a" shows input extract from normal (WT) Py infected BMK cells; lane "b" shows cell extract from lane "a" pulled down with GST alone; lane "c" shows cell extract from lane "a" pulled down with GST-mSal2 fusion protein. Lanes "d" to "h" show pulldown assays using cell extracts of 3T3 cells transfected with WT large T antigen or TMD25 large T antigen cDNA: lane "d" shows the input extract from 3T3 transfected with WT large T antigen cDNA; lane "e" shows the input extract from 3T3 transfected with TMD25 large T antigen cDNA; lane "f" shows the extract of WT large T antigen cDNA transfected 3T3 cells pulled down with GST alone; lane "g" shows the extract of WT large T antigen cDNA transfected 3T3 cells pulled down with GST-mSal2 fusion protein; and lane "h" shows the extract of TMD25 large T antigen cDNA transfected 3T3 cells pulled down with GST-mSal2 fusion protein. Normal large T antigen synthesized during infection of BMK efficiently binds the GST-mSal2

fragment (lanes a to c). Comparing extracts of 3T3 cells transfected with either wild-type, or TMD-25, large T antigen cDNAs only the wild-type shows binding (lanes d to g).

To confirm the large T-p150^{sal2} interaction *in vivo*, 3T3 cells were doubly transfected with a vector expressing full length GST-mSal2 and either wild-type, or TMD-25 mutant, large T antigen cDNAs (Fig. 4B Left Panel). Cell extracts were pulled down with glutathione beads. After electrophoresis and transfer, the filter was blotted with anti-large T antigen antibody to show the binding of wild-type or mutant large T antigen. The same filter was blotted again with a monoclonal antibody against mSal2 to show that the level of expression of GST-mSal2 is similar in both the wild-type large T antigen and the TMD25 large T antigen experiments. Each lane is labeled and the input equaled 3% of the extracts used in the co-precipitation assay. Complexes containing normal large T antigen were readily recovered, but no evidence of binding was seen with the mutant large T antigen.

A further experiment was done to confirm the interaction between the large T protein and p150^{sal2} during a lytic viral infection. An extract of wild-type virus-infected BMK cells was prepared 24 hours post-infection and incubated with polyclonal serum made against the amino-terminal mSal2 fragment. The anti-mSal2 immunoprecipitate was separated and blotted with an anti-T monoclonal antibody. A portion of the large T antigen present in the virus-infected cell extract clearly immunoprecipitated with mSal2, showing that these two proteins interact (Fig. 4B Right Panel). Polyoma large T and p150^{sal2} most likely interact directly through their carboxyl-terminal regions, although additional factors may be involved in mediating the binding.

Example 4: TMD-25 Is Defective In Virus Growth And Tumor Induction In The Newborn Mouse

Newborn mice were inoculated with either wild-type or TMD-25 mutant virus and followed for development of tumors. The ability of TMD-25 to replicate and spread in the newborn mouse was examined by whole mouse

section hybridization (Dubensky et al., *J. Virol.* 65:342-349 (1991). At ten days post inoculation the mutant showed no signs of replication and spread while the wild-type virus established a disseminated infection with extensive replication in many tissues (Fig. 5A).

Tests for virus replication were carried out on ten-day old animals by whole mouse section hybridization using a ^{35}S-labelled viral DNA probe (Fig. 5A). Newborn mice were inoculated subcutaneously with TMD25 or PTA (1 X 10^6 each) and sacrificed ten days later. Frozen sections were probed with ^{35}S labeled viral DNA with overnight exposure. Wild-type PTA showed strong replication in kidney, skin, and bones, while the TMD25 mutant showed no sign of viral replication in any of the organs. Table 2 shows a comparison of tumor induction profile between mSal2 binding mutant TMD25 and wild-type PTA viruses. Newborn mice were inoculated as described above, and sacrificed five months later. Pathological examinations were performed for tumor profile. Wild-type virus rapidly established a disseminated infection and induced a broad spectrum of tumors (Table 2). In contrast, TMD-25 failed to replicate and spread. The only tumors found in mutant-infected mice were subcutaneous fibrosarcomas and these developed only at the site of virus inoculation. Since TMD-25 is defective in replication but retains normal middle and small T functions, these findings are consistent with the expectation that the input mutant virus would be able to infect and transform cells locally but be unable to spread and induce a broad spectrum of tumors.

Direct tests of the mutant's transforming ability were carried out using standard assays with an established line of rat embryo fibroblasts (Dahl et al., *Mol. Cell Biol.* 16:2728-2735 (1996)). Transformation of these cells does not depend on virus replication, and middle T alone suffices for transformation (Raptis et al., *Mol. Cell Biol.* 5:2476-2485 (1985)). Mutant virus-infected cells gave rise to foci resembling those induced by wild-type virus; cells derived from one such focus were confirmed, by DNA sequencing, to carry the mutant viral genome. Using DNA transfection followed by measuring colony formation in soft agar, transforming efficiencies were found to be essentially identical for

wild-type and mutant viral DNAs – approximately 10-20 colonies/10^5 cell/μg viral DNA. The failure of TMD-25 to induce tumors at sites distant from the site of inoculation is therefore not due to any defect in transforming ability, but rather to its inability to replicate and establish a disseminated infection.

To investigate whether binding of $p150^{sal2}$ by large T antigen is necessary for viral DNA replication, low molecular weight DNA from BMK cells infected by wild-type or mutant virus was extracted and analyzed by Southern hybridization. The results show clearly that the mutant was unable to replicate its DNA in the non-permissive host (BMK) cells 36 hr post infection (Fig. 5B, Left Panel). BMK cells were infected with TMD25 and wild-type virus (Wt Py). Low molecular weight DNA was isolated at 0, 18, 36 hrs post infection (p.i.) for Southern blot with virus DNA probe. These results suggest that $p150^{sal2}$ can act, directly or indirectly, to inhibit viral DNA replication.

Furthermore, when over expressed in normal 3T3 cells, $p150^{sal2}$ inhibited wild-type viral DNA replication in a dose-dependent manner (Fig. 5B, Right Panel). Polyoma origin clone pUCori (Ori) and large T –expressing plasmid, (Wt LT cDNA), were cotransfected with increasing amount of plasmid expressing mSal2. Newly replicated DNA was detected with origin specific probe (top). The filter was striped and re-probed with LT and origin specific probe to show that similar amount of origin and LT DNA were present in each transfection. These results show that $p150^{sal2}$ imposes a block to viral DNA replication and that the block can be overcome by wild-type large T antigen.

Example 5: Expression Pattern Of $p150^{sal2}$ In The Mouse

Normal mouse tissues were extracted and tested for expression of $p150^{sal2}$ by Western blot (Fig. 6). Tissues from ten to twelve-day old mice were dissected and extracted in NP-40 lysis buffer. 200 μg of protein from various tissues were loaded onto each lane as labeled. The proteins were detected using a monoclonal antibody against the amino-terminus of mSal2. Tissue from brain, kidney, lung, bladder, and uterus clearly shows expression of the protein, while tissue from liver, skeletal muscle, spleen, salivary gland, and heart was either negative or low

in expression. These results are consistent with those reported earlier by Northern analysis. The finding that the kidney and lung are sites of strong expression is also consistent with the natural history of transmission of polyoma, which is thought to infect through the lung and amplify primarily in the kidney. Successful growth in these tissues would require the virus to be able to overcome any block to replication imposed by mSal2. TMD-25 fails to replicate its DNA in normal mouse cells, and overexpression of mSal2 blocks normal viral DNA replication.

Example 6: Expression Of hSal2 In Human Ovarian Tumors

The *hSal2* gene has been mapped to chromosome 14q12 but was not recognized initially as a tumor suppressor gene. It was subsequently shown by others that this region of 14q is associated with a loss of homozygosity in 49% of ovarian cancers (Bandera et al., *supra*) and about 25 % of bladder cancers (Chang et al., *supra*). These findings, along with the underlying rationale of 'tumor host range' selection, suggest the possibility that *sal2* may function as a tumor suppressor. To test this possibility more directly, a screen for $p150^{sal2}$ expression was carried out on extracts of ovarian carcinomas, the results of which are summarized in Fig. 7, a Western blot of human ovarian tumors. The expression level of $p150^{sal2}$ in 20 ovarian carcinomas was compared with that of normal ovarian epithelial cells (N) in two panels. Fifty micrograms of protein were loaded in each lane and blotted with polyclonal antibody against $p150^{sal2}$. Each ovarian carcinoma was labeled by its case number. Arrows indicate the normal position of p150. A polyclonal anti-p150 antibody made against the mouse protein clearly recognizes the human protein (Fig. 3B above). A band of the same apparent molecular weight is seen in extracts of normal human ovarian epithelial cells ('HOSE').

In situ staining with anti-p150 was carried out on frozen sections of normal ovary and several ovarian carcinomas, as well as in human 293 cells. Fig. 8 shows expression of $p150^{sal2}$ in human 293 cells. A polyclonal antibody, HM867, raised against mSal2 carboxyl-terminus, was used to detect human $p150^{sal2}$ in human 293 cells (lane +). As a negative control, the same protein

extract was blotted with HM867 antibody that had first been depleted by incubation with the same antigen used to raise it (lane -). As a further example of $p150^{sal2}$ expression, Fig. 9 shows immunostaining of $p150^{sal2}$ in the human ovary and in ovarian tumors. Fig. 9A shows immunostaining of normal human ovarian tissue with a polyclonal serum preadsorbed with mSal2 protein. In the left-hand panel, normal human ovarian tissue is stained with a polyclonal serum preadsorbed with $p150^{sal2}$. In the right-hand panel, normal ovarian tissue is stained with polyclonal serum against $p150^{sal2}$. Fig. 9B shows six ovarian carcinoma tissue samples that were stained for $p150^{sal2}$ (c thru h), where "T" stands for tumor cells and "S" stands for stromal cells. The insert in "h" shows cytoplasmic staining for $p150^{sal2}$. The nuclear staining of normal epithelial cells is readily apparent, but in the ovarian tumor cells the staining is reduced or cytoplasmic.

Example 7: A Point Mutation, S73C, In Human Sal2 Is Present In Some Ovarian Tumors.

DNAs from twenty-one ovarian carcinomas were digested and analysed by Southern hybridization using a probe of *hSal2* coding sequences. *hSal1* sequences were used as an unlinked internal control. No evidence of loss or gross rearrangement of the *hSal2* locus was seen in any of the tumors examined. However, deletions of 1kb or less would not have been detected. The absence of $p150^{sal2}$ expression in a majority of ovarian cancers may reflect mechanisms other than loss of the *hSal2* gene itself, such as silencing of expression through promoter methylation, alterations in an upstream regulatory factor, or factors leading to instability of the protein itself.

To test for small mutations, DNAs from four tumors were extracted and the entire *hSal2* coding regions sequenced on both strands. Two tumors from the panel shown in Fig. 7 that were positive for $p150^{sal2}$ expression and two that were negative were chosen. The two negative tumors 327 and 523 showed no changes when compared to the controls and all showed sequences identical to the published genomic sequence (Genbank AE000658 and AE000521; Boysen et al, *Genome Res.* 330:330-338 (1997)). The two $p150^{sal2}$–positive tumors each

showed a cysteine (TGT) substitution for serine (TCT) at position 73 (position 73 of SEQ ID NO:1), based on the first methionine in exon 1a (Kohlhase et al., *Mamm Genome* 11:64-69 (2000). The sequencing results showed only TGT in tumor 432 and a mixture of TGT and TCT in tumor 528. The serine codon TCT has been found at this position in all normal DNAs sequenced thus far (Kohlhase et al., *Genomics* 38:291-298 (1996); Boysen et al., *Genome Res.* 330:330-338 (1997)), indicating that '73S' is a frequent normal allele. To know whether the S73C substitution represents a somatic mutation or germ line polymorphism, normal DNA from case 432 was sequenced. The result showed only TGT at codon 73, indicating that the *hSal2* allele encoding cysteine represents a germ-line polymorphism in this individual. DNAs from six ovarian carcinoma cell lines were also sequenced and one showed the same S73C substitution as seen in case 432 and another a G744R substitution.

An example of the loss of the 73S allele is shown in Fig. 11. For this experiment, DNA was isolated from matched normal and ovarian tumor tissues. The 73S and 73C alleles were distinguished by PCR amplification and subsequent Mbo II digestion of a 318 bp product covering the region containing amino acid 73. In addition to a common Mbo II site (used to monitor the digestion status), this region contains another Mbo II site for the 73S allele, but not for the 73C allele (this is the discriminating Mbo II recognition site). Complete digestion of 73S allele by Mbo II produced three fragments (171 bp, 94 bp and 53 bp) while 73C allele produced two fragments (256 bp and 53 bp fragments-indicated by arrows). These fragments were resolved by electrophoresis on a 2% agarose gel. Although it is difficult to avoid the existence of normal tissue in the tumor used to isolate DNA, the intensity of the 73S bands (171 bp and 94 bp) is largely reduced indicating the loss of 73S allele (patient number 1). In this figure, "U" indicates undigested amplification product, "S" indicates a 73S homozygote control, "C" indicates a 73C homozygote control, and "S/C" indicates a 73S/C heterozygote control. The respective identification number of ovarian tumor patients is shown on top of their matched normal "N" and tumor "T" DNA.

Example 8: mSal2 Suppresses Growth of Ovarian Carcinoma Cells

To characterize the biological function of Sal2, the ovarian carcinoma cell line SKOV3 was transfected with an *mSal2* expression vector. SKOV3 cells were transfected with pcDNA-mSal2 (P150) or pcDNA3 vector (Mock), incubated in 0.5% serum for 48 hours, then in 15% serum and 100 μM BrdU for 20 hours. This cell line expresses little or no $p150^{sal2}$ as is indicated by Western analysis. Cells were examined by BrdU incorporation for DNA synthesis, for $p150^{sal2}$ expression, and for DAPI staining (Fig. 10A). The percent of cells in S-phase decreased from 57% in the control to 19% in cells expressing $p150^{sal2}$. In addition, 30-50% of cells expressing $p150^{sal2}$ appeared to be apoptotic as judged by DAPI staining compared to less than 10% of control cells. Arrows in frame 1 of Fig. 10A indicate a cell expressing $p150^{sal2}$ that is BrdU-negative. Arrows in frame 2 of Fig. 10A indicate an apoptotic cell expressing $p150^{sal2}$ with fragmented nuclear bodies as shown in the merged image. The bar graph in Fig. 10A shows the percentage of BrdU-positive cells in Mock and $p150^{sal2}$ expressing cells. In a colony reduction assay conducted over 14 days, a clear reduction in viable SKOV3 cells was seen in cells transfected with the expression vector, reflecting both growth suppressive, and apoptosis inducing activity of $p150^{sal2}$ (Fig. 10B). Similar efficiencies of transfection (approximately 20%) were confirmed by a co-transfected GFP expression plasmid.

Example 9: Experimental Procedures

Selection of tumor host-range mutants

The cell lines used as permissive hosts include TCMK-1 (Black et al., *Proc. Soc. Exper. Biol. Med.* 114:721-727 (1963)) purchased from ATCC) and A6241 (Lukacher et al., *J. Exp. Med.* 181:1683-1692 (1995); Velupillai et al., *J. Virology* 73: 10079-10085 (1999)). Primary baby mouse kidney cells (BMK) were used as the non-permissive host. The genome of polyoma virus strain PTA was digested at the single BamHI site and cloned into pBlueScript (Stratagene) to create PTAHI. PTAHI was amplified in the Mut D strain of *E. coli* (Schaaper et

al., *Proc. Natl. Acad. Sci. U.S.A.* 85:8126-8130 (1998)) to accumulate mutations randomly throughout the viral genome.

Yeast two-hybrid screening

The polyoma PTA large T antigen carboxyl-terminal fragment (amino acids 333-781) was cloned into pGBT9 (Clontech) to generate pGBT9ITC used as a "bait" to screen a 9.5 to 10.5 day-old whole mouse embryo cDNA library in pVP16 (Vojtek et al., *Cell* 75:205-214 (1993)). Transformation and selection were performed according to the recommendations from Clontech.

Generation of TMD25 with a minimum deletion

Large T antigen carboxyl-terminal deletions used in the yeast two-hybrid analysis were generated on pGBT9ITC using the Transformer site-directed *in vitro* mutagenesis kit (Promega) according to manufacturer's recommendations.

Cloning of full length mSal2 cDNA

A complete cDNA sequence for *mSal2* was obtained by RACE (Frohman) using Marathon cDNA amplification kit (Clontech) and RT-PCR products from BMK cells.

RFLP Test to Identify a Polymorphism in Sal2

Amino acid 73 of human $p150^{sal2}$ is polymorphic. This amino acid may be a serine encoded by the codon TCT (73S) or a cysteine encoded by the codon TGT (73C). The two alleles may be distinguished by PCR amplification of the genomic region encompassing the sequence encoding hSal2 amino acid 73 and digesting the PCR product using either the restriction enzyme Mob II or Ear I. These enzymes cut the DNA close to the codon encoding amino acid 73. The primers used to amplify the DNA prior to digestion with Mob II were, 5'-CTTGTTAATTAGAGCCTCGGTATACC-3' (SEQ ID NO:7) and 5'-GCACGGAGGACCCAGAATCTGG-3' (SEQ ID NO:8).

The PCR cycle used was 98°C for 2 minutes followed by 35 cycles of 94°C for 1 minute, 55°C for 1 minute, and 68°C for 1 minute. After the last PCR

cycle, the reaction was incubated at 72°C for 10 minutes. The PCR products were digested with Mob II in a solution containing 5 µl PCR mixture, 2 µl enzyme buffer (10 fold concentrated), 12 µl water, and 1 µl Mob II (5 units/µl). The restriction digest was performed at 37°C for two hours followed by heating the reaction to 70°C for twenty minutes prior to loading ten to twenty microliters of the mixture onto a 2% agarose gel. Five microliters of undigested PCR product were added to a control lane on the gel. The expected size of the uncut PCR product is 318 bp. The expected Mob II restriction fragments for the 73S allele are 171, 94, and 53 bp and the expected Mob II restriction fragments for the 73C allele are 265 and 53 bp. A mixture of the 73S and 73C alleles would be expected to yield fragments of 265, 171, 94, and 53 bp. The 53 bp fragment is common to both alleles and may be used to monitor the digestion status in order to distinguish between heterozygotes and an incomplete digestion.

In vitro GST pull-down assay

Full-length polyoma normal large T antigen cDNA and TMD25 large T antigen cDNA were cloned into pcDNA3 to create CMVLT and CMVTMDLT respectively. The mSal2 fragment (amino acids 841-971), containing the last zinc finger pair, was cloned into pGEX4T1 (Pharmacia) to generate GST-mSal2 fusion protein in *E. coli*. The fusion protein was bound to glutathione-Sepharose 4B beads (purchased from Pharmacia) according to the manufacturer's instructions. For the association of GST-mSal2 fusion with large T antigen, BMK cells infected by PTA, or 3T3 cells transfected with wild-type or TMD25 large T antigen expression constructs CMVLT or CMVTMDLT, were extracted with NP-40 lysis buffer (pH 7.9) (Benjamin et al., *Proc. Natl. Acad. Sci. U.S.A.* 67:394-399 (1970)). 500µl of cell lysate were incubated with 50µl of 50% GST-Sal2 or GST beads for 2 hours. After washing four times with PBS, the bound protein was subjected to Western blot analysis using monoclonal antibody F4, which recognizes T antigens (Dahl et al., *Mol. Cell. Biol.* 16:2728-2735 (1996)).

In vivo GST pull-down assay

The full-length mSal2 coding region was cloned into a eukaryotic GST fusion vector, pEBG (Luo et al., *J. Biol. Chem.* 270:23681-23687 (1995)) to generate the construct pEBGSAL. NIH 3T3 cells were co-transfected with pEBGSAL and CMVLT or CMVTNDLT in a ratio of 1 to 1 using Lipofect2000 (Gibco/BRL) according to the manufacturer's protocol. The cells were harvested 48 hours post transfection. The lysate was centrifuged at 3,000 rpm and the supernatant was incubated with 50-100 µl glutathione-Sepharose 4B beads for 2 hours. The beads were washed four times with PBS containing 0.01% NP-40 and the bound proteins were immunoblotted with the F4 antibody and an antibody against $p150^{sal2}$ (Dahl et al., *supra* (1996)).

In vivo Co-immunoprecipitation of mSal2 and Polyoma Large T

Fifty microliters of 50% protein A beads (Pharmacia) were incubated with purified rabbit polyclonal anti-amino-terminal mSal2 antibody or normal rabbit IgG in 1 ml NP-40 lysis buffer at 4°C for 2 hours, followed by washing the beads four times with PBS. Two milligrams of total protein, made from BMK cells infected with wild-type virus, were extracted 24 hours post infection and incubated with either the anti-mSal2 or normal lgG beads in NP-40 lysis buffer containing 1% BSA for 2 hours at 4°C. After the incubation, the beads were washed four times with 0.1% Tween-20 in PBS and the proteins were separated by SDS-PAGE. Polyoma large T and mSal2 were detected using anti-T and anti-mSal2 monoclonal antibodies.

Viral DNA Replication Assays

Plasmid pUCori and the polyoma origin replication assay are described in Gjorup et al. (*Proc. Natl. Acad. Sci.* USA 91:12125-12129 (1994)). Cells were grown on 6 well plates and infected with virus or transfected with DNA. Low molecular weight DNA was isolated as described by Hirt (*J. Mol. Biol.* 26:365-369 (1967)). After purification, the DNAs were resuspended in 80 µl of water. One to five micrograms of DNA were subjected to restriction digestion. For

virus infection experiments, the viral genome was first linearized with Eco RI. For transfection experiments, pUCori and CMVLT were first digested with Dpn I and Hind III. The newly synthesized pUCori DNA is Dpn I resistant because of the lack of methylation in eukaryotic cells and the input plasmid DNA is sensitive to Dpn I digestion because of the *E. coli* methylation of the recognition site. The DNA fragment was resolved on a 1% agarose gel for Southern analysis using origin specific and LT specific probes.

Western blots for detection of $p150^{sal2}$

Tissue extracts were prepared from C3H/BiDa mice by homogenization in NP-40 lysis buffer (pH 7.9) and centrifugation at 8,000 rpm. Fifty micrograms of protein (Bio-Rad Assay) from each sample was separated by SDS-PAGE and blotted on nitrocellulose membranes. A monoclonal antibody against mSal2 was used to detect $p150^{sal2}$.

Stripping Western Filters for Reprobing

After first antibody probing, the used filter is incubated in stripping solution (50 mM Tris-Cl, pH 6.8, 2% SDS and 100 mM β-mercaptoethanol) for 30 minutes at 60°C. The filter is washed twice in PBS and tested for the absence of the previously used antibody by development and exposure to an X-ray film. This procedure ensures that the filter can be used again in subsequent Western analyses.

Analysis of ovarian carcinomas

Surgical samples of human ovarian tissue were obtained under a protocol approved by the Human Subjects Committee of the Brigham and Women's Hospital. Ovarian tumor tissues were pulverized in liquid nitrogen and lysed in a buffer (1% Triton X-100, 21 μg/ml aprotinin, 1 mM phenylmethylsulfonyl fluoride, 0.5 μg/ml leupeptin, 4.9 mM $MgCl_2$, and 1mM vanadate in PBS). The MicroBAC Protein Assay Kit (Pierce) was used for protein quantitation.

Twenty-five micrograms of protein from each sample were separated on an SDS-polyacrylamide gel and blotted on nitrocellulose membranes. A rabbit polyclonal antiserum that cross-reacts with hSal2 was used to detect $p150^{sal2}$. Specifically, this antiserum was raised against a GST-mouse $p150^{sal2}$ fusion protein that was first purified using Affinity Pak Immobilized Protein A (Pierce) according to manufacturer's instructions followed by an incubation with GST saturated glutathione beads (Pharmacia) in PBS for 30 minutes to eliminate antibodies against GST. As a negative control, the purified antibody was preadsorbed with the GST- $p150^{sal2}$ fusion protein.

Frozen sections of normal or tumor samples were fixed in Neutral Formalin for 10 minutes and permeabilized in cold ethanol/acidic acid (3:1) for 15 min. After washing four times in PBS for 10 minutes each, the sections were antibody stained and processed using Vectastain ABC kit (Vector Laboratories) following the manufacturer's instructions.

DNAs were extracted from human ovarian carcinomas and from primary cultures of ovarian epithelial cells obtained by scraping the surface of normal ovarian tissue. DNA from normal human foreskin was used as a control. The coding region with the 0.4 kb intron of *hSal2* was amplified using the primer pair (5'-CCACAACCATGGCGAATCCGAG-3') (SEQ ID NO:5) and (5'-GGTGATGGAAGGCGAACAGCCAGG-3') (SEQ ID NO:6). Long range PCR was performed (98°C 4 min, then 94°C 1 min, 60°C 1 min, 68°C 4 min, for 35 cycles) and sequencing was carried out using the High Throughput Core of the Dana Farber-Harvard Cancer Center. The coding region was sequenced twice and additional sequencing of both strands was performed for regions with suspected mutations. The resulting sequence was compared with the published *hSal2* cDNA sequence and genomic sequence.

BrdU Incorporation

SKOV3 cells were transfected with pcDNA-*mSal* and the pcDNA 3 vector using BRL Lipofectamine 2000 according to the manufacturer's recommendations. Five to seven hours post transfection the cells were fed with

0.5% calf serum. After 48 hours, the cells were incubated with a medium containing 15% calf serum with 100 mM BrdU for 20 to 24 hours. A monoclonal antibody against BrdU (Amersham) was used to detect the incorporation. The cells were fixed, permeabilised and stained according to Amersham's recommendations except that a purified rabbit polyclonal antibody against the mSal2 carboxyl-terminus was mixed with the BrdU antibody for the detection of both BrdU incorporation and $p150^{sal2}$ expression. Secondary antibodies (anti-mouse Rhodamine and anti-rabbit Oregon Green) were also mixed. Cells were examined under fluorescence microscopy in order to identify BrdU and $p150^{sal2}$ positive cells.

Colony Reduction Assay

SKOV3 cells were transfected with a pcDNA-*mSal* or a pcDNA3 vector in a 6 well plates using 2 μg of DNA each. To monitor the transfection efficiency, 0.5 μg of pEGFPN1 (Clontech) was added to the test DNA in a separate tube. Transfection was performed according to GIBCO/BRL's recommendations using LIPOFECTAMIN 2000. Twenty-four hours after the transfection, the cells were re-seeded in 10 cm plates with medium containing 600 μg/ml G418 (GIBCO/BRL) and 10% calf serum. The EGFP expression was also monitored at this time. The G418 containing medium (neomycin medium) was changed every 3 to 4 days until mock-transfected cells had died and neomycin resistant colonies became apparent.

Preparation of DNA for microinjection

As but one example, DNA clones for microinjection are prepared by cleaving the DNA with enzymes appropriate for removing the bacterial plasmid sequences and subjecting the DNA fragments to electrophoresis on 1% agarose gels in TBE buffer (Sambrook et al. (1989)). The DNA bands are visualized by staining with ethidium bromide and the band containing the desired DNA sequences is excised. The excised band is then placed in dialysis bags containing 0.3 M sodium acetate pH 7.0. The DNA is electroeluted into the dialysis bags,

extracted with phenol/chloroform (1:1), and precipitated by the addition of two volumes of ethanol. The DNA is then redissolved in 1 ml of low salt buffer (0.2 M NaCl, 20 mM Tris, pH 7.4, and 1 mM EDTA) and purified on an Elutip-D.TM (Schleicher and Schuell) column. The column is first primed with 3 ml of high salt buffer (1M NaCl, 20 mM Tris, pH 7.4, and 1 mM EDTA) followed by washing with 5 ml of low salt buffer. The DNA solutions are passed through the column three times to bind the DNA to the column matrix. After one wash with 3 mls of low salt buffer, the DNA is eluted with 0.4 ml of high salt buffer and precipitated by the addition of two volumes of ethanol. DNA concentrations are measured by absorption at 260 nm in a UV spectrophotometer. For microinjection, DNA concentrations are adjusted to 5 μg/ml in 5 mM Tris, pH 7.4 and 0.1 mM EDTA. Other methods for purification of DNA for microinjection are also described in Hogan et al. (*Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1986)); in Palmiter et al. (*Nature* 300:611 (1982)); in the Qiagenologist, *Application Protocols*, 3rd edition, published by Qiagen, Inc., Chatsworth, Calif.; and in Sambrook et al. (*Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989)). The procedures for manipulation of the rodent embryo and for microinjection of DNA are described in detail in Hogan et al. (*Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., (1986)), the teachings of which are incorporated herein.

Animal experiments

Whole mouse section hybridizations (Dubensky et al., *J. Virol.* 68:342-349 (1991)) and tumor profiles (Dawe et al., *Am. J. Pathol.* 127:243-261 (1987)) were performed as described in these publications.

Production of transgenic mice and rats

The following is but one preferred means of producing transgenic mice. This general protocol may be modified by those skilled in the art.

Female mice six weeks of age are induced to superovulate with a 5 IU injection (0.1 cc, IP) of pregnant mare serum gonadotropin (PMSG; Sigma) followed 48 hours later by a 5 IU injection (0.1 cc, IP) of human chorionic gonadotropin (hCG, Sigma). Females are placed together with males immediately after hCG injection. Twenty-one hours after hCG injection, the mated females are sacrificed by CO_2 asphyxiation or cervical dislocation and embryos are recovered from excised oviducts and placed in Dulbecco's phosphate buffered saline with 0.5% bovine serum albumin (BSA, Sigma). Surrounding cumulus cells are removed with hyaluronidase (1 mg/ml). Pronuclear embryos are then washed and placed in Earle's balanced salt solution containing 0.5% BSA (EBSS) in a 37.5°C incubator with humidified atmosphere at 5% CO_2, 95% air until the time of injection. Embryos can be implanted at the two-cell stage.

Randomly cycling adult female mice are paired with vasectomized males. Swiss Webster or other comparable strains can be used for this purpose. Recipient females are mated at the same time as donor females. At the time of embryo transfer, the recipient females are anesthetized with an intraperitoneal injection of 0.015 ml of 2.5% avertin per gram of body weight. The oviducts are exposed by a single midline dorsal incision. An incision is then made through the body wall directly over the oviduct. The ovarian bursa is then torn with watchmakers forceps. Embryos to be transferred are placed in DPBS (Dulbecco's phosphate buffered saline) and in the tip of a transfer pipet (about 10 to 12 embryos). The pipet tip is inserted into the infundibulum and the embryos are transferred. After the transferring the embryos, the incision is closed by two sutures.

The preferred procedure for generating transgenic rats is similar to that described above for mice (Hammer et al., *Cell* 63:1099-112 (1990). For example, thirty-day old female rats are given a subcutaneous injection of 20 IU of PMSG (0.1 cc) and 48 hours later each female placed with a proven, fertile male. At the same time, 40-80 day old females are placed in cages with vasectomized males. These will provide the foster mothers for embryo transfer.

The next morning females are checked for vaginal plugs. Females who have mated with vasectomized males are held aside until the time of transfer. Donor females that have mated are sacrificed (CO_2 asphyxiation) and their oviducts removed, placed in DPBA (Dulbecco's phosphate buffered saline) with 0.5% BSA and the embryos collected. Cumulus cells surrounding the embryos are removed with hyaluronidase (1 mg/ml). The embryos are then washed and placed in EBSs (Earle's balanced salt solution) containing 0.5% BSA in a 37.5°C incubator until the time of microinjection.

Once the embryos are injected, the live embryos are moved to DPBS for transfer into foster mothers. The foster mothers are anesthetized with ketamine (40 mg/kg, IP) and xulazine (5 mg/kg, IP). A dorsal midline incision is made through the skin and the ovary and oviduct are exposed by an incision through the muscle layer directly over the ovary. The ovarian bursa is torn, the embryos are picked up into the transfer pipet, and the tip of the transfer pipet is inserted into the infundibulum. Approximately 10 to 12 embryos are transferred into each rat oviduct through the infundibulum. The incision is then closed with sutures, and the foster mothers are housed singly.

Generation of Knockout Mice

The following is but one example for the generation of a knockout mouse and the protocol may be readily adapted or modified by those skilled in the art.

Embryonic stem cells (ES), for example, 10^7 AB1 cells, may be electroporated with 25 μg targeting construct in 0.9 ml PBS using a Bio-Rad Gene Pulser (500 μF, 230 V). The cells may then be plated on one or two 10-cm plates containing a monolayer of irradiated STO feeder cells. Twenty-four hours later, they may be subjected to G418 selection (350 μg/ml, Gibco) for 9 days. Resistant clones may then be analyzed by Southern blotting after *Hind* III digestion, using a probe specific to the targeting construct. Positive clones are expanded and injected into C57BL/6 blastocysts. Male chimeras may be back-crossed to C57BL/6 females. Heterozygotes may be identified by Southern blotting and intercrossed to generate homozygotes.

In addition, knockout mice may also be generated by site-specific recombination methods using, for example, the FLP/FRT system or the Cre-lox system. These systems are described in the specification as well as in, for example, U.S. Patent Number 5,527,695, Lyznik et al. (*Nucleic Acid Research* 24:3784-3789 (1996), and Kilby et al. (Trends in Genetics 9:413-421 (1993)).

The targeting construct used in making the knockout animal may result in the disruption of the gene of interest, e.g., by insertion of a heterologous sequence containing stop codons, or the construct may be used to replace the wild-type gene with an altered form of the same gene, e.g., a mutant *Sal2* gene. In addition, the targeting construct may contain a sequence that allows for conditional expression of the gene of interest. For example, a sequence may be inserted into the gene of interest that results in the protein not being expressed in the presence of tetracycline. Such conditional expression of a gene is described in, for example, WO 94/29442, WO 96/40892, WO 96/01313, and Yamamoto et al. (*Cell* 101:57-66 (2000).

Table 2. Tumor profiles of mutant TMD-25 and wild-type PTA virus

	TMD-25	PTA[1]
Fraction of mice with tumors	7/7	32/32
Mean age at necropsy	202d.	82d.
Epithelial tumors:		
Hair follicle	0/7	32/32
Thymus	0/7	29/32
Mammary gland	0/7	16/32
Salivary gland	0/7	23/32
Mesenchymal tumors:		
Fibrosarcomas	7/7[2]	1/32
Renal medulla	0/0	7/32
Bone	0/0	6/32

1. Data on PTA is taken from Dawe et al. (*Am. J. Pathol.* 127:243-261, 1987).

2. Subcutaneous fibrosarcomas were found only at the site of virus inoculation.

Table 3. Summary of $p150^{sal2}$ expression in human ovarian carcinomas

$p150^{sal2}$ Status	Number of Cases	Percent
7		
Positive	6	30
Negative	10	50
Altered*	4	20

* Refers to the apparent size of the Sal2 protein, which is different from that of normal ovarian epithelial cells.

Other Embodiments

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known or customary practice within the art to which the invention pertains and may be applied to the essential features hereinbefore set forth.

All references cited herein are hereby incorporated by reference.

We claim:

CLAIMS

1. A method of identifying a cellular protein involved in the susceptibility to proliferative disease, said method comprising the steps of:

a) infecting a normal cell and an abnormally proliferating cell with a collection of uncharacterized mutant viruses;

b) identifying a mutant virus from the collection that can grow in said abnormally proliferating cell and can not grow in said normal cell; and

c) identifying the mutated viral gene or mutated protein in said virus,

which allows said virus to grow on said abnormally proliferating cell; and

d) screening to identify the cellular protein which interacts with the wild-type viral protein, but not said mutated viral protein.

2. The method of claim 1, wherein said abnormally proliferating cell is uncharacterized.

3. The method of claim 1, further comprising identifying a cellular protein that can interact with a wild-type viral protein that corresponds to said mutant viral protein, wherein said cellular protein is not a retinoblastoma tumor suppressor protein.

4. The method of claim 3, wherein the step of identifying said cellular protein comprises using an assay that detects protein-protein interactions.

5. The method of claim 4, wherein said assay is a GST-pulldown assay.

6. The method of claim 3, further comprising isolating a gene encoding said cellular protein.

7. The method of claim 1, wherein said virus has a mammalian host range.

8. The method of claim 7, wherein said mammal is a human.

9. The method of claim 1, wherein said virus is selected from the group consisting of simian virus 40 virus, human polyoma virus, parnovirus, papilloma virus, herpes virus, and primate adenoviruses.

10. The method of claim 1, wherein said cellular protein is a tumor suppressor protein.

11. The method of claim 1, wherein said cellular protein is a proto-oncogene product.

12. A tumor host range virus isolated using the method of claim 1.

13. A method of determining the presence or absence of an alteration in the genetic material of a cell, said method comprising determining whether a cell can act as a permissive host for the propagation of a characterized T-HR mutant, said T-HR mutant being capable of propagating in an abnormally proliferating cell and not being capable of propagating in a normal cell, wherein said characterized T-HR mutant is unable to propagate in a cell carrying a mutation in the retinoblastoma or p53 gene.

14. The method of claim 13, wherein the presence of said genetic alteration is indicative of an organism carrying this genetic alteration being at an increased risk of developing a proliferative disease.

15. The method of claim 13, wherein said alteration in the genetic material is in a tumor suppressor gene.

16. The method of claim 13, wherein said alteration in the genetic material is in a proto-oncogene.

17. The method of claim 13, wherein said characterized T-HR mutant has been characterized as being complemented by a mutation in a specific tumor suppressor gene or proto-oncogene, wherein said tumor suppressor or proto-oncogene are not the retinoblastoma or p53 gene.

18. The method of claim 13, wherein said cell is a cell from a mammal.

19. The method of claim 18, wherein said mammal is a human.

20. A method of killing an abnormally proliferating cell comprising the steps of:

(i) contacting an abnormally proliferating cell with a T-HR mutant; and

(ii) allowing said T-HR mutant to lyse said cell.

21. The method of claim 20, wherein said abnormally proliferating cell is a mammalian cell.

22. The method of claim 21, wherein said mammalian cell is a human cell.

23. The method of claim 20, wherein said abnormally proliferating cell is in a mammal with a proliferative disorder.

24. The method of claim 23, wherein said mammal is a human.

25. The method of claim 20, wherein said T-HR mutant is administered in a pharmaceutically acceptable carrier.

26. The method of claim 20, wherein said T-HR mutant is administered by a method selected from the group consisting of parenteral, intravenous, intraperitoneal, intramuscular, subcutaneous, and subdermal injection.

27. The method of claim 20, wherein said T-HR mutant is administered by a method selected from the group consisting of orally, nasally, topically, and as an aerosol.

28. The method of claim 20, wherein said TH-R mutant is selected from the group consisting of, simian virus 40, human polyoma virus, herpes virus, primate adenoviruses, parnovirus, and papilloma virus.

29. The method of claim 20, wherein said T-HR mutant is specific for a cell carrying a *Sal2* mutation.

30. The method of claim 29, wherein said T-HR mutant is the TMD-25 T-HR mutant virus.

31. A method of identifying a mammal having or at increased risk of acquiring a proliferative disease, said method comprising the step of determining whether there is a proliferative disease-associated alteration in a *Sal2* nucleic acid of said mammal.

32. The method of claim 31, wherein said method is for identifying a mammal having a proliferative disease.

33. The method of claim 31, wherein said method is for identifying a mammal at increased risk of acquiring a proliferative disease.

34. The method of claim 31, wherein said mammal is a human.

35. The method of claim 34, wherein said proliferative disease-associated alteration comprises the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

36. The method of claim 31, wherein said determining is done by polymerase chain reaction (PCR) amplification, single nucleotide polymorphism (SNP) determination, restriction fragment length polymorphism (RFLP) analysis, hybridization analysis, or mismatch detection analysis.

37. The method of claim 31, wherein said method comprises the steps of:

(i) contacting a first nucleic acid probe which is specific for binding to

said human *Sal2* nucleic acid containing said alteration with a nucleic acid from a cell from said mammal under conditions which allow said first nucleic acid probe to anneal to complementary sequences in said cell; and

(ii) detecting duplex formation between said first nucleic acid probe and said complementary sequences.

38. The method of claim 37, wherein said first nucleic acid probe is derived from the human *Sal2* nucleic acid containing a proliferative disease-associated alteration.

39. The method of claim 37, further comprising a second nucleic acid probe, wherein said first and second nucleic acid probes are PCR primers, and wherein said human *Sal2* nucleic acid or a fragment thereof is amplified using PCR between steps (i) and (ii).

40. The method of claim 37, wherein said cell is from a physiological sample containing abnormally proliferating tissue.

41. The method of claim 37, wherein said cell is from a physiological sample of normal tissue.

42. The method of claim 37, wherein said alteration comprises the substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

43. A method of identifying a mammal having or at increased risk of acquiring a proliferative disease, said method comprising the step of determining whether there is a proliferative disease-associated alteration in a Sal2 protein of said mammal.

44. The method of claim 43, wherein said method is for identifying a mammal having a proliferative disease.

45. The method of claim 43, wherein said method is for identifying a mammal at increased risk of acquiring a proliferative disease.

46. The method of claim 43, wherein said mammal is a human.

47. The method of claim 43, wherein said method comprises the use of an antibody specific for a human Sal2 protein.

48. The method of claim 47, wherein said antibody comprises an antibody specific for a proliferative disease-associated mutant Sal2 protein.

49. A knockout mouse comprising a knockout mutation in a genomic *mSal2* gene.

50. The knockout mouse of claim 49, wherein said mouse further comprises a nucleic acid construct including a mutant *Sal2* gene.

51. The knockout mouse of claim 40, wherein said mutant *Sal2* gene is conditionally expressed.

52. The knockout mouse of claim 40, wherein said mutant *Sal2* gene encodes a protein that contains a substitution of a Cys for the Ser at position 73 of SEQ ID NO:1.

53. A transgenic mouse whose genome comprises a nucleic acid construct
including a *Sal2* nucleic acid, which is operably linked to transcriptional regulatory elements and encodes a Sal2 protein.

54. The transgenic mouse of claim 53, wherein said Sal2 protein is mutant.

55. The transgenic mouse of claim 53, wherein said transcriptional regulatory elements include a promoter that is a tissue-specific promoter.

56. The transgenic mouse of claim 55, wherein said nucleic acid is expressed such that the protein is produced at detectable levels in cells selected from the group consisting of ovarian, bladder, and colon cells.

57. The transgenic mouse of claim 55, wherein said transcriptional regulatory elements include a promoter that is an ovary-specific promoter.

58. The transgenic mouse of claim 53, wherein said Sal2 nucleic acid is a human Sal2 nucleic acid.

59. The transgenic mouse of claim 54, wherein said mouse develops ovarian tumors.

60. The transgenic mouse of claim 59, wherein said ovarian tumors metastasize.

61. A cell line derived from cells isolated from said transgenic mouse of claim 53.

62. A method of identifying a compound which alters cell proliferation, said method comprising:

a) contacting a first cell with a test compound, and

b) measuring whether said test compound alters proliferation in said first cell, relative to a second cell not contacted with said test compound, wherein said first and second cells have a proliferative disease-associated alteration in a Sal2 nucleic acid.

63. The method of claim 62, wherein the ability of said test compound to alter proliferation is measured by measuring the ability of a virus to propagate in said first cell contacted with said test compound, relative to said second cell not contacted with said test compound.

64. The method of claim 63, wherein said virus is a T-HR mutant virus.

65. The method of claim 62, wherein said first and second cells are mammalian cells.

66. The method of claim 62, wherein said first and second cells are in the same mammal or in different mammals.

67. The method of claim 66, wherein said mammal is a transgenic mouse.

68. The method of claim 66, wherein said mammal is a knockout mouse comprising a knockout mutation in a genomic *mSal2* gene.

69. The method of claim 62, wherein said first and second cells are ovarian cells.

70. A method of identifying a compound which alters cell proliferation, the method comprising:

a) exposing a cell or a cell extract to a test compound, and

b) measuring whether said test compound alters Sal2 levels, relative to Sal2 levels in a cell or cell extract not exposed to said test compound.

71. The method of claim 70, wherein said Sal2 is Sal2 protein.

72. The method of claim 70, wherein said Sal2 is *Sal2* nucleic acid.

73. The method of claim 70, wherein said measuring is by measuring Sal2 protein levels.

74. The method of claim 70, wherein said measuring is by measuring *Sal2* nucleic acid levels.

75. The method of claim 70, wherein said cell has a proliferative disease-associated alteration in a *Sal2* nucleic acid or said extract is from a cell having a proliferative disease-associated alteration in a *Sal2* nucleic acid.

76. The method of claim 70, wherein said exposing is with a cell and said cell is an ovarian cell.

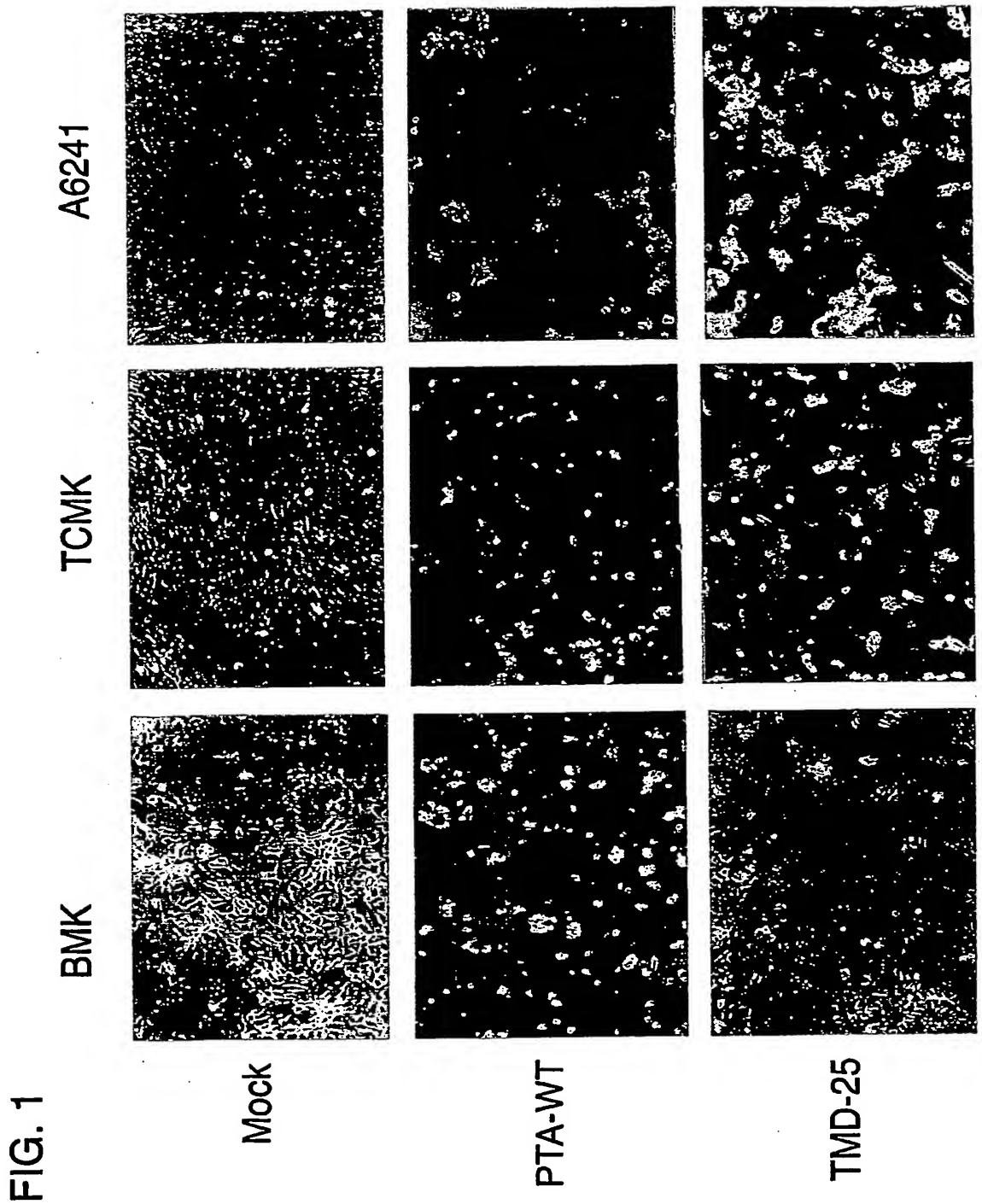


FIG. 1

FIG. 2A

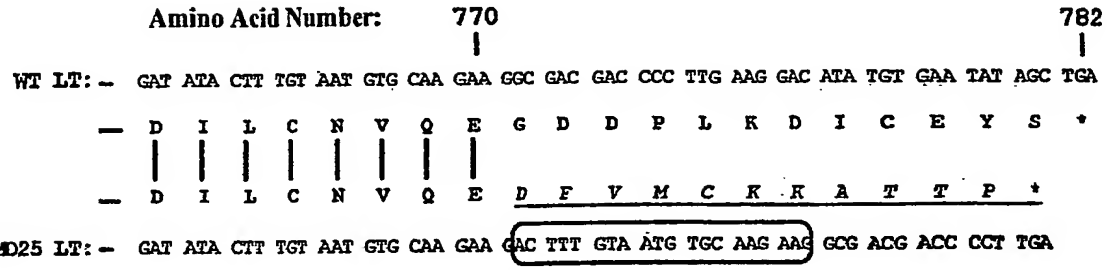


FIG. 2B

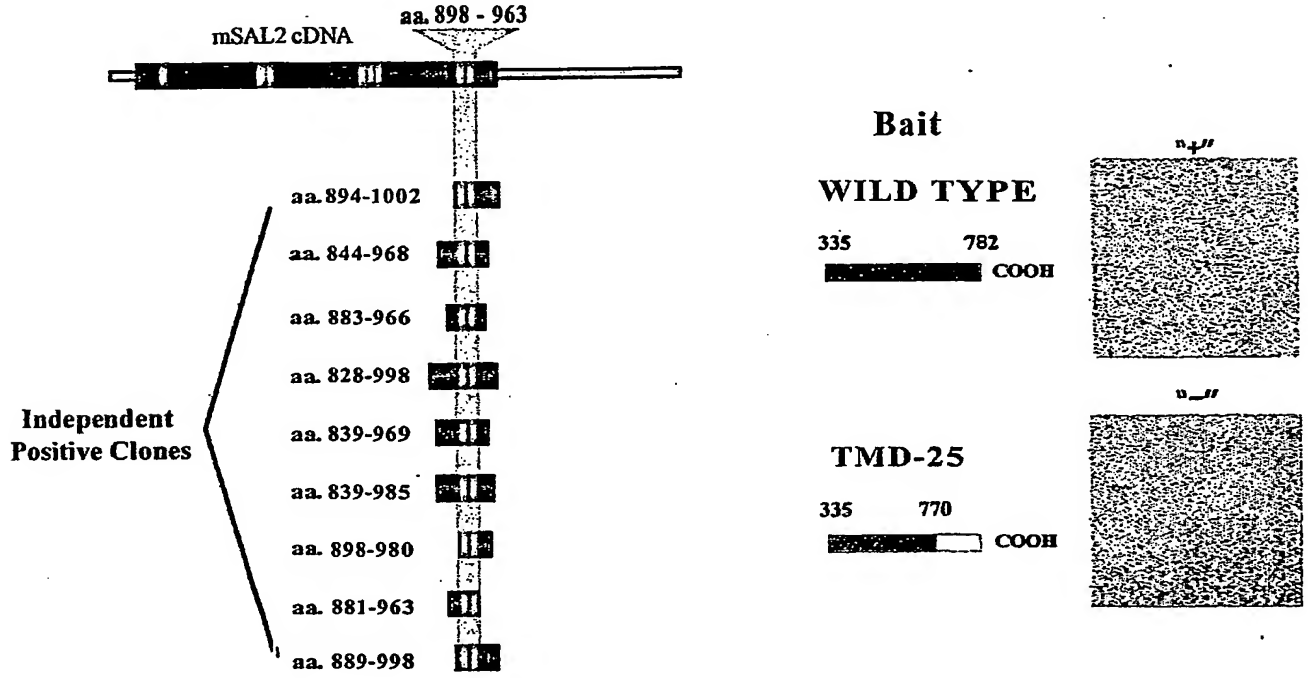


FIG. 2C

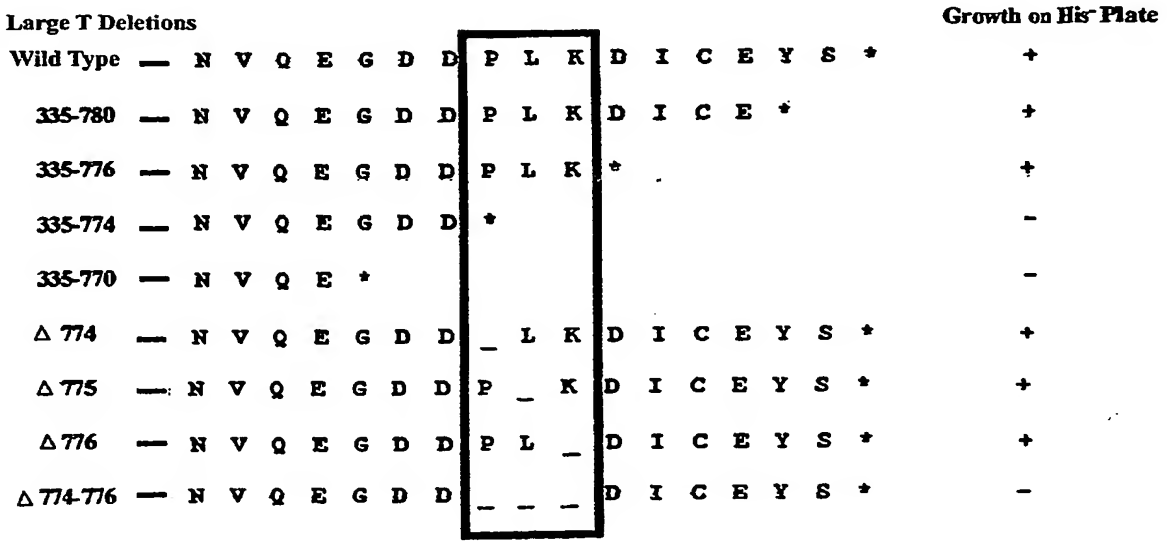


FIG. 3A

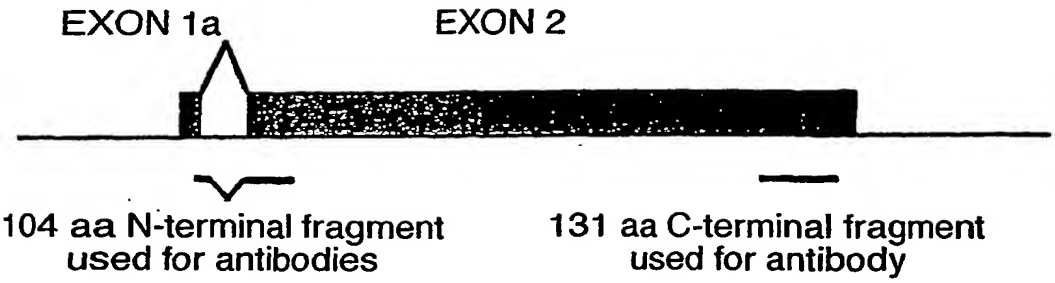


FIG. 3B

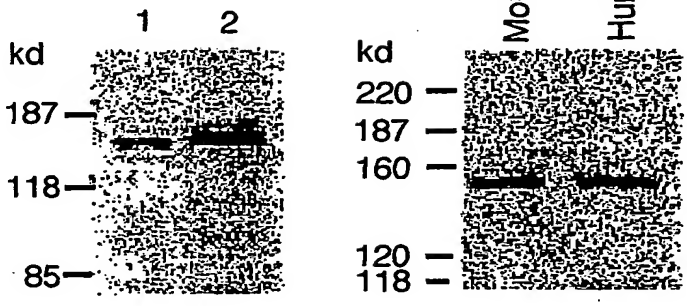


FIG. 3C

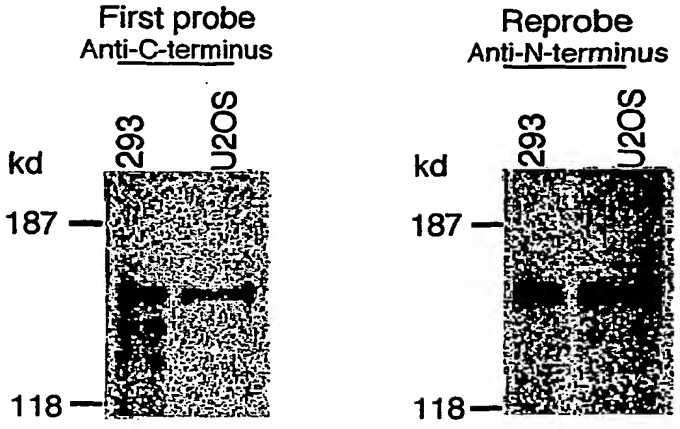


FIG. 4A

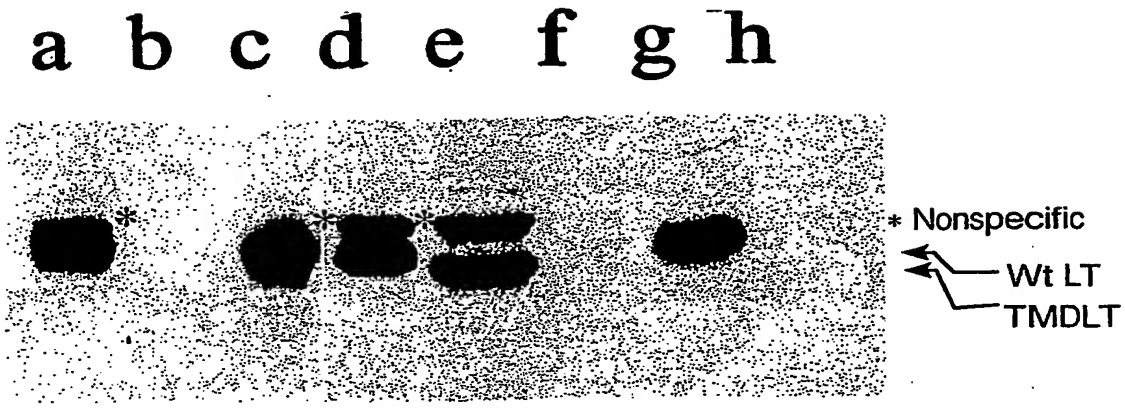


FIG. 4B

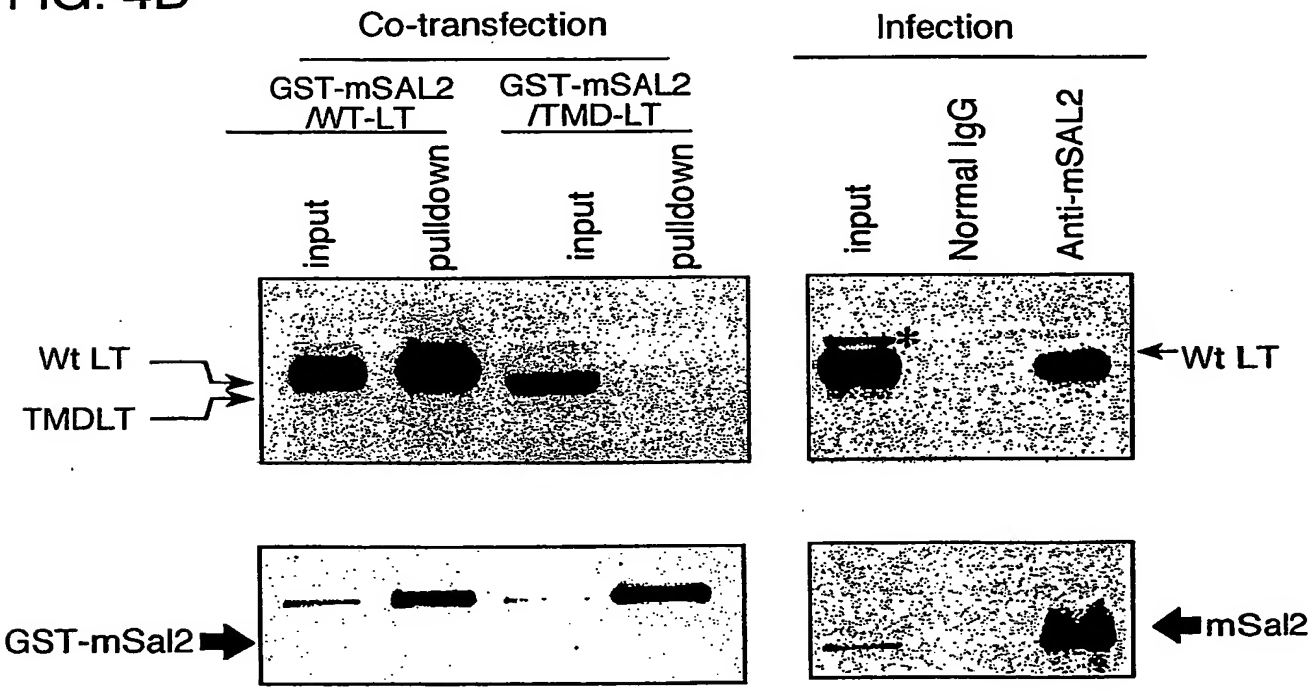
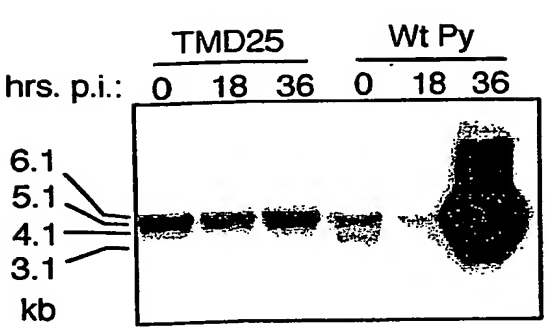


FIG. 5A

TMD-25 WILD TYPE

FIG. 5B



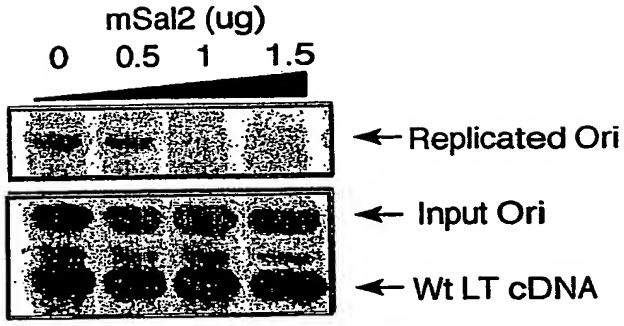


FIG. 6

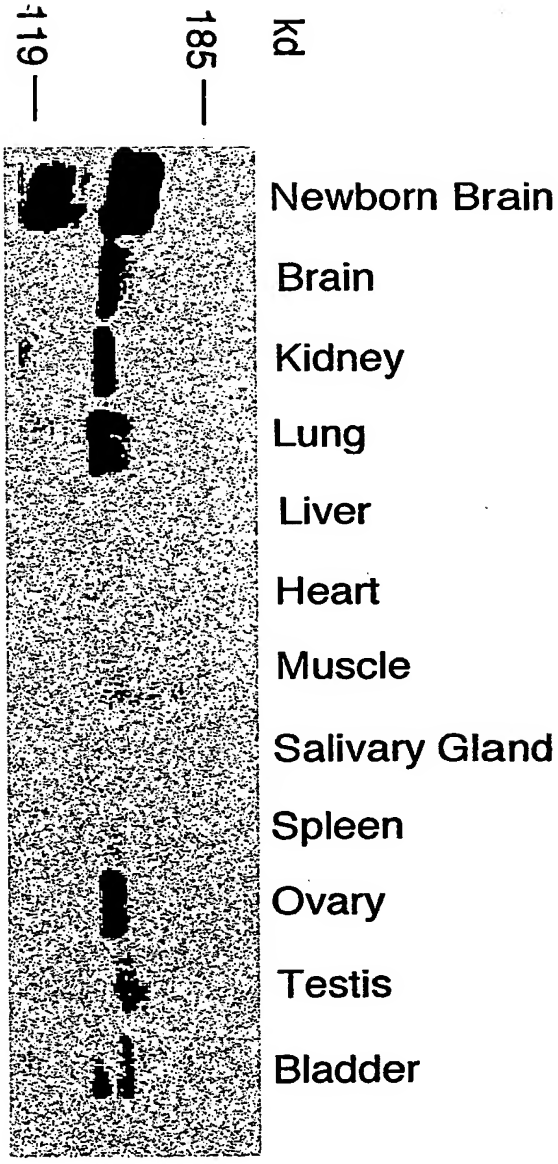


FIG. 7

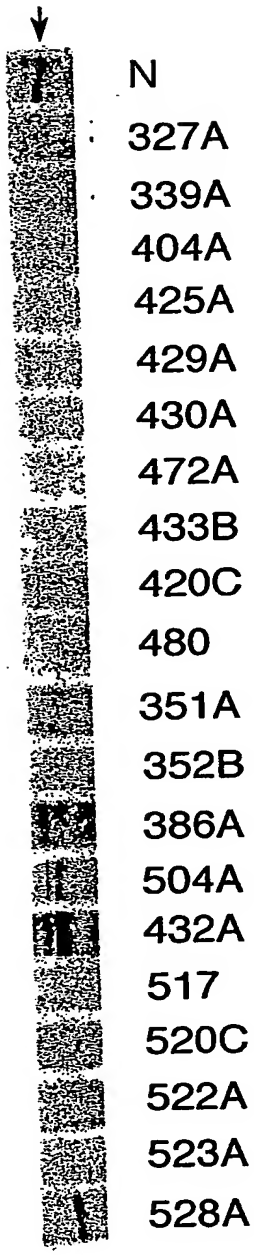
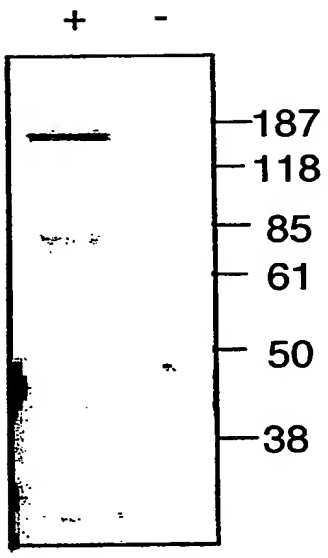

N
327A
339A
404A
425A
429A
430A
472A
433B
420C
480
351A
352B
386A
504A
432A
517
520C
522A
523A
528A

FIG . 8



293 Cell extract

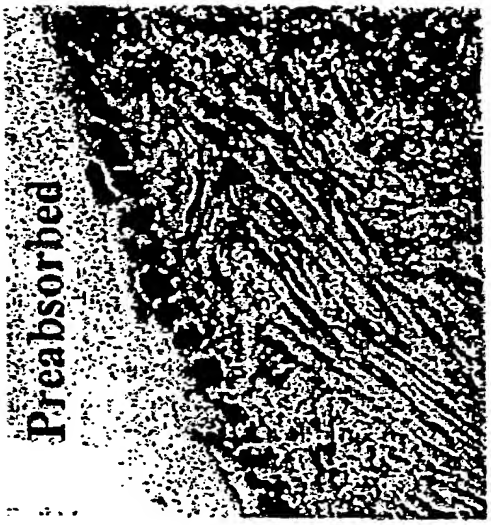


FIG. 9A

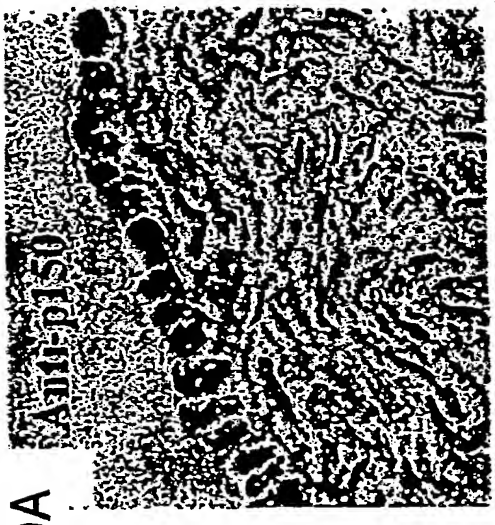


FIG. 9B

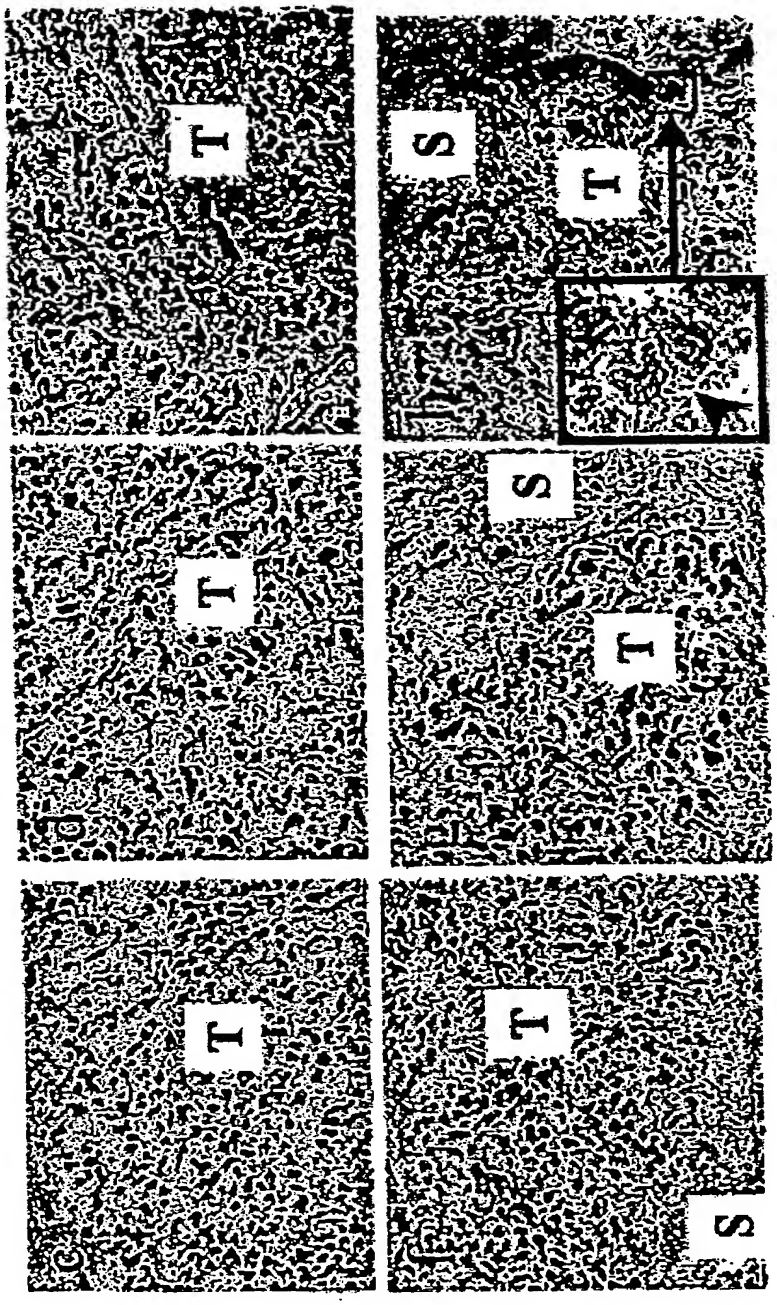
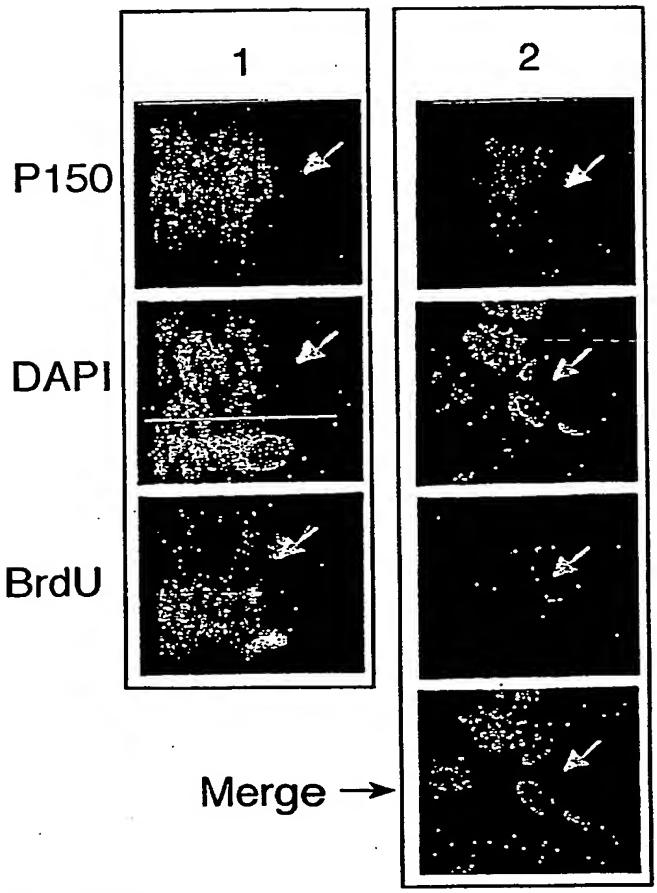


FIG. 10A



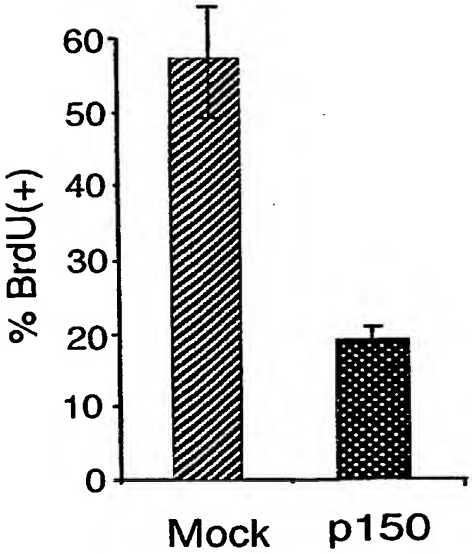


FIG. 10B

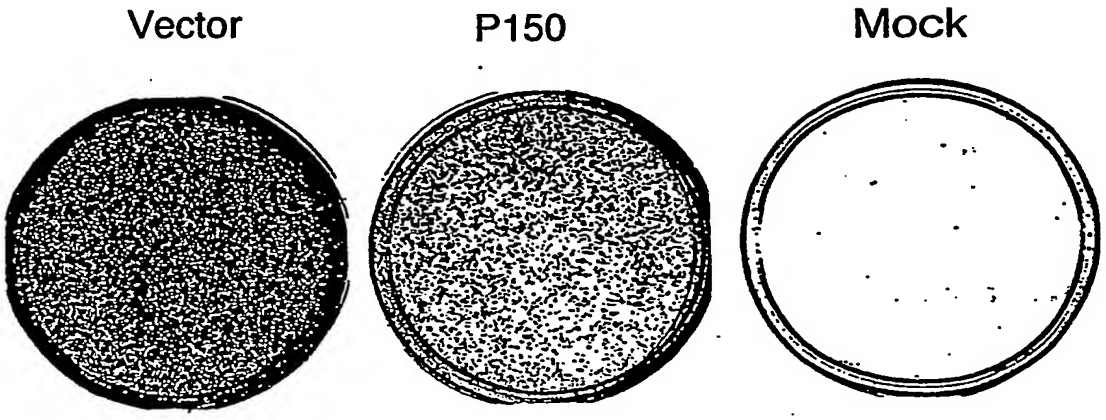
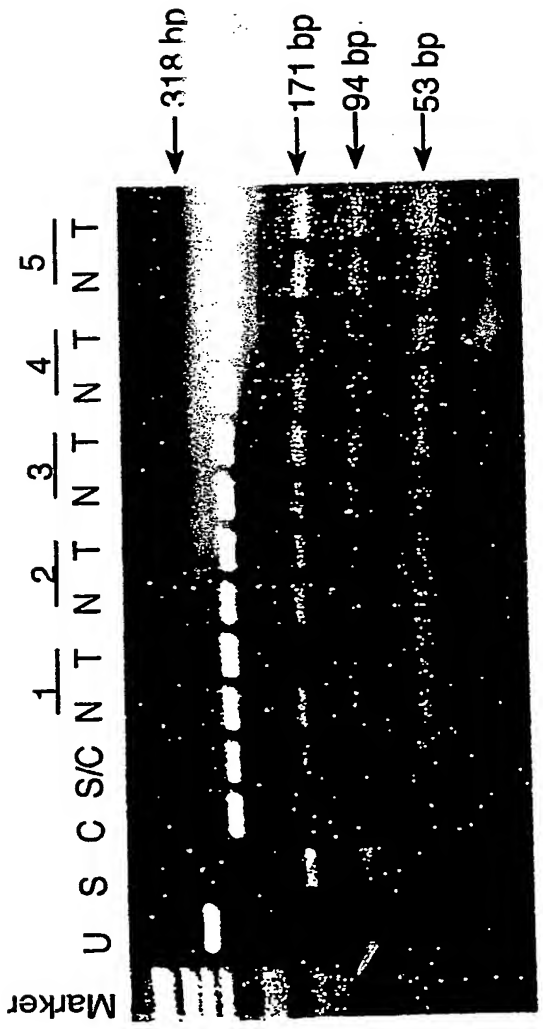


FIG. 11



SEQUENCE LISTING

<110> President and Fellows of Harvard College
The Brigham and Women's Hospital, Inc.

<120> Diagnosing and Treating Cancer Cells
Using Mutant Viruses

<130> 00742/062WO2

<150> 60/216,723
<151> 2000-07-07

<150> 09/812,633
<151> 2001-03-19

<150> 09/812,471
<151> 2001-03-19

<160> 21

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1005
<212> PRT
<213> Homo Sapiens

<400> 1

```
Met Ala His Glu Ser Glu Arg Ser Ser Arg Leu Gly Val Pro Ala Gly
 1               5                  10                  15
Glu Pro Ala Glu Leu Gly Gly Asp Ala Ser Glu Glu Asp His Pro Gln
            20                  25                  30
Val Cys Ala Lys Cys Cys Ala Gln Phe Thr Asp Pro Thr Glu Phe Leu
        35                  40                  45
Ala His Gln Asn Ala Cys Ser Thr Asp Pro Pro Val Met Val Ile Ile
    50                  55                  60
Gly Gly Gln Glu Asn Pro Asn Asn Ser Ser Ala Ser Ser Glu Pro Arg
65                  70                  75                  80
Pro Glu Gly His Asn Asn Pro Gln Val Met Asp Thr Glu His Ser Asn
                85                  90                  95
Pro Pro Asp Ser Gly Ser Ser Val Pro Thr Asp Pro Thr Trp Gly Pro
            100                 105                 110
Glu Arg Arg Gly Glu Glu Ser Ser Gly His Phe Leu Val Ala Ala Thr
        115                 120                 125
Gly Thr Ala Ala Gly Gly Gly Gly Gly Leu Ile Leu Ala Ser Pro Lys
    130                 135                 140
Leu Gly Ala Thr Pro Leu Pro Pro Glu Ser Thr Pro Ala Pro Pro Pro
145                 150                 155                 160
Pro Pro Pro Pro Pro Pro Pro Pro Gly Val Gly Ser Gly His Leu Asn
                165                 170                 175
Ile Pro Leu Ile Leu Glu Glu Leu Arg Val Leu Gln Gln Arg Gln Ile
            180                 185                 190
```

```
His Gln Met Gln Met Thr Glu Gln Ile Cys Arg Gln Val Leu Leu Leu
        195                 200                 205
Gly Ser Leu Gly Gln Thr Val Gly Ala Pro Ala Ser Pro Ser Glu Leu
    210                 215                 220
Pro Gly Thr Gly Thr Ala Ser Ser Thr Lys Pro Leu Leu Pro Leu Phe
225                 230                 235                 240
Ser Pro Ile Lys Pro Val Gln Thr Ser Lys Thr Leu Ala Ser Ser Ser
                245                 250                 255
Ser Ser Ser Ser Ser Ser Ser Gly Ala Glu Thr Pro Lys Gln Ala Phe
            260                 265                 270
Phe His Leu Tyr His Pro Leu Gly Ser Gln His Pro Phe Ser Ala Gly
        275                 280                 285
Gly Val Gly Arg Ser His Lys Pro Thr Pro Ala Pro Ser Pro Ala Leu
    290                 295                 300
Pro Gly Ser Thr Asp Gln Leu Ile Ala Ser Pro His Leu Ala Phe Pro
305                 310                 315                 320
Ser Thr Thr Gly Leu Leu Ala Ala Gln Cys Leu Gly Ala Ala Arg Gly
                325                 330                 335
Leu Glu Ala Thr Ala Ser Pro Gly Leu Leu Lys Pro Lys Asn Gly Ser
            340                 345                 350
Gly Glu Leu Ser Tyr Gly Glu Val Met Gly Pro Leu Glu Lys Pro Gly
        355                 360                 365
Gly Arg His Lys Cys Arg Phe Cys Ala Lys Val Phe Gly Ser Asp Ser
    370                 375                 380
Ala Leu Gln Ile His Leu Arg Ser His Thr Gly Glu Arg Pro Tyr Lys
385                 390                 395                 400
Cys Asn Val Cys Gly Asn Arg Phe Thr Thr Arg Gly Asn Leu Lys Val
                405                 410                 415
His Phe His Arg His Arg Glu Lys Tyr Pro His Val Gln Met Asn Pro
            420                 425                 430
His Pro Val Pro Glu His Leu Asp Tyr Val Ile Thr Ser Ser Gly Leu
        435                 440                 445
Pro Tyr Gly Met Ser Val Pro Pro Glu Lys Ala Glu Glu Glu Ala Ala
    450                 455                 460
Thr Pro Gly Gly Gly Val Glu Arg Lys Pro Leu Val Ala Ser Thr Thr
465                 470                 475                 480
Ala Leu Ser Ala Thr Glu Ser Leu Thr Leu Leu Ser Thr Ser Ala Gly
                485                 490                 495
Thr Ala Thr Ala Pro Gly Leu Pro Ala Phe Asn Lys Phe Val Leu Met
            500                 505                 510
Lys Ala Val Glu Pro Lys Asn Lys Ala Asp Glu Asn Thr Pro Pro Gly
        515                 520                 525
Ser Glu Gly Ser Ala Ile Ser Gly Val Ala Glu Ser Ser Thr Ala Thr
    530                 535                 540
Leu Met Gln Leu Ser Lys Leu Met Thr Ser Leu Pro Ser Trp Ala Leu
545                 550                 555                 560
Leu Thr Asn His Phe Lys Ser Thr Gly Ser Phe Pro Leu Pro Leu Cys
                565                 570                 575
Ala Arg Ala Leu Gly Ala Ser Pro Ser Glu Thr Ser Lys Leu Gln Gln
            580                 585                 590
Leu Val Glu Lys Ile Asp Arg Gln Gly Ala Val Ala Val Thr Ser Ala
        595                 600                 605
Ala Ser Gly Ala Pro Thr Thr Ser Ala Pro Ala Pro Ser Ser Ser Ala
    610                 615                 620
Ser Ser Gly Pro Asn Gln Cys Val Ile Cys Leu Arg Val Leu Ser Cys
```

```
625                 630                 635                 640
Pro Arg Ala Leu Arg Leu His Tyr Gly Gln His Gly Gly Glu Arg Pro
                645                 650                 655
Phe Lys Cys Lys Val Cys Gly Arg Ala Phe Ser Thr Arg Gly Asn Leu
            660                 665                 670
Arg Ala His Phe Val Gly His Lys Ala Ser Pro Ala Ala Arg Ala Gln
        675                 680                 685
Asn Ser Cys Pro Ile Cys Gln Lys Lys Phe Thr Asn Ala Val Thr Leu
    690                 695                 700
Gln Gln His Val Arg Met His Leu Gly Gly Gln Ile Pro Asn Gly Gly
705                 710                 715                 720
Thr Ala Leu Pro Glu Gly Gly Gly Ala Ala Gln Glu Asn Gly Ser Glu
                725                 730                 735
Gln Ser Thr Val Ser Gly Ala Gly Ser Phe Pro Gln Gln Gln Ser Gln
            740                 745                 750
Gln Pro Ser Pro Glu Glu Glu Leu Ser Glu Glu Glu Glu Glu Glu Asp
        755                 760                 765
Glu Glu Glu Glu Glu Asp Val Thr Asp Glu Asp Ser Leu Ala Gly Arg
    770                 775                 780
Gly Ser Glu Ser Gly Gly Glu Lys Ala Ile Ser Val Arg Gly Asp Ser
785                 790                 795                 800
Glu Glu Ala Ser Gly Ala Glu Glu Glu Val Gly Thr Val Ala Ala Ala
                805                 810                 815
Ala Thr Ala Gly Lys Glu Met Asp Ser Asn Glu Lys Thr Thr Gln Gln
            820                 825                 830
Ser Ser Leu Pro Pro Pro Pro Pro Pro Asp Ser Leu Asp Gln Pro Gln
        835                 840                 845
Pro Met Glu Gln Gly Ser Ser Gly Val Leu Gly Gly Lys Glu Glu Gly
    850                 855                 860
Gly Lys Pro Glu Arg Ser Ser Ser Pro Ala Ser Ala Leu Thr Pro Glu
865                 870                 875                 880
Gly Glu Ala Thr Ser Val Thr Leu Val Glu Glu Leu Ser Leu Gln Glu
                885                 890                 895
Ala Met Arg Lys Glu Pro Gly Glu Ser Ser Ser Arg Lys Ala Cys Glu
            900                 905                 910
Val Cys Gly Gln Ala Phe Pro Ser Gln Ala Ala Leu Glu Glu His Gln
        915                 920                 925
Lys Thr His Pro Lys Glu Gly Pro Leu Phe Thr Cys Val Phe Cys Arg
    930                 935                 940
Gln Gly Phe Leu Glu Arg Ala Thr Leu Lys Lys His Met Leu Leu Ala
945                 950                 955                 960
His His Gln Val Gln Pro Phe Ala Pro His Gly Pro Gln Asn Ile Ala
                965                 970                 975
Ala Leu Ser Leu Val Pro Gly Cys Ser Pro Ser Ile Thr Ser Thr Gly
            980                 985                 990
Leu Ser Pro Phe Pro Arg Lys Asp Asp Pro Thr Ile Pro
        995                 1000                1005
```

<210> 2
<211> 16080
<212> DNA
<213> Homo sapiens

<400> 2

atatcacacc ccagctggct atgtaatcat gaaataagga gaaacacata aatatttggt 60
taaaacacct ttaatgatag agggaaagac actaatatct cccgtctgtt cttgacattt 120
tactaggtta ggaagctctg gagcctacag cttgaggaga agccatcgtt caagtcagtc 180
aatagcaaaa ccctcactct ctcctcctca gaactcctgt tccaaatgat cctatgttaa 240
gagtaaatac tacaactcat tacaagacgg agaggcaggg aggacgccac ctggagctgg 300
gactcttaag aaccagacaa tgacaaagac acaagcccca gcctacggat aggcaaaatg 360
ggtaggggtc ttgaaagagg aagataagga aaatacaagg ggccagggaa taaaggaggg 420
agttatctaa aactagaagc atactagtgc taggaaatcc cccatgatcc ctggtacacc 480
tctgcacact atgtcactat tagcccaaaa gaatattaac gagaatgtcc acattcacaa 540
gaatttgagg ccttttccct tacatcatgt ccctttctta gtcacatagg taccagcaag 600
ccctatgttc tagcaacatt ccttaactct ctcatcatta gttcatcaac catgctgacc 660
aaaaatgctc cttaaagata cgaacttcac atttcccaaa tatctcctgg gagacctctt 720
ggcaagaaat cagcttgttt cccaactttg agaggtcatc atgaatgaga agctggagag 780
gtcttggcac actgaccagc caaaaccttt accttaatgt gaccatcagg ggatttactg 840
ggaaaatttt cctatgccct tccttcattt ctccctactt cctagggttg ggtcaccaat 900
tactggagca tcttcagtac cggcaccttc tggagcaggg ggaggaagaa ggaatgtaca 960
gtttgctact tcttgtctat gatgggcttc tcaggcactg ccttgggtgc aggaggctga 1020
aataggaggg gggctgtctt ctccttggct tccctggatc ccattgttgg aggcaccttc 1080
ccagccacag ttcctaggcc aaacagcact ggtggggcca ggcttggagt ggtagtggag 1140
gtggagctgg aattccaggg cttcatgggc aggccatttg acaggaatgc cacatactgg 1200
ttctagaaag ataggggacc catacccacc agctgagcag aaaggtcacc ccagaggagt 1260
ggcactgggc cctccagaga cagctgccag ccctttttgg ctaggctgca atgccaaatg 1320
taggtgctca ggtgcaccta ccaaagggaa agggagagga gagaggaggg ggaagaaggg 1380
tcacaccagg gaagctggag agggttcccc ttgagaaagc tgcagagaat ctatgttcct 1440
caggtacaaa gaatgaggag ggaagaaaaa ttccttaggg ggccatcccc ttgtaagcac 1500
agtaatttcc aagctcaggg actacagaaa agccactagg gacataacat gttaagaact 1560
tagagaaaaa gacaaaatca gggctcataa ctctgggagg tccttttgtg aagctgtttc 1620
tgctctgtgg gacaaagagc agcaggtaca gaaaaacagg ctcatgggat cgtggggtca 1680
tcttttcggg gaaaggggga gagccctgtg gaggtgatgg aaggcgaaca gccagggact 1740
agagaaagag cagcaatatt ctgagggcca tgggggggcaa agggctgtac ctggtggtgt 1800
gccaggagca tatgcttctt gagggtagcc cgctcaagaa agccctgcct gcagaaaaca 1860
caagtgaaga gcggcccctc cttggggtgg gtcttctgat gctcctccag agctgcctgg 1920
gagggaaagg cctggccaca cacttcgcag gcctttctgc tgctgctctc tcctggctcc 1980
tttctcattg cctcctgcag gctcagctcc tctaccaagg tcacgctggt ggcttccgct 2040
tctggggtga gtgctgatgc cggacttgag cttctctccg gtttgccccc ctcttccttg 2100
cctcctaaaa caccactgct tccctgctcc attggctgag gctgatccag gctgtcaggt 2160
ggtggtggtg gtggcaaaga agactgttga gtagttttct cattactgtc catctccttc 2220
ccagctgtgg ctgctgccgc cactgtcccc acctcctcct ctgccccaga tgcctcttct 2280
gaatcacctc tcactgatat tgccttctca cctccactct ctgagcctct ccctgccagg 2340
gaatcttcat cagtcacatc ttcctcttct tcctcatcct cctcttcctc ctcctcagac 2400
aactcctctt ccggtgatgg ctgctgggac tgctgctggg ggaaactccc tgccccggag 2460
actgtagatt gctcggagcc attctcctga gcagctcctc caccttcagg gagtgcagta 2520
ccaccgttgg ggatctggcc ccccaggtgc atccggacat gctgctgcag agtgacagca 2580
ttggtgaact tcttctggca gatggggcag gaattctgtg cccgggcagc tggactggcc 2640
ttgtggccca cgaaatgtgc acgcagatta cccctggtgg agaaggctct gccacacact 2700
ttgcatttga agggcctctc acctccatgt tggccataat gaaggcgtag ggcccgagga 2760
cagctaagca ctcggagaca gatgacacac tggttaggtc cagaagaggc tgaggatgaa 2820
ggtgcagggg cagaggtggt gggggctcct gaggcagctg aggtcaccgc cacagctcct 2880
tgccggtcaa tcttttctac cagttgctgc agctttgatg tctcagaggg tgaggccccc 2940
aagggctcta gcacataggg gaaggggaag ctgccagtgg acttgaagtg gttggtaagc 3000
agtgcccagc ttggtagtga agtcaccaac ttacttagtt gcatgcgagt tgccgtgcta 3060
ctttctgcca ctccactgat ggctgagccc tcactccctg ggggggtgtt ttcatcagct 3120
ttattcttgg gttccactgc tttcatgagc acaaacttat tgaaagcagg gagtcctgga 3180
gccgtggctg tgcctgcact ggtggagagc agagtcaggc tctctgtggc actgagtgct 3240
gttgtggagg ccaccagagg cttgcgctca acccctccac ctggagtggc tgcctcctcc 3300

tcggccttct ctggtggcac ggacatacca taaggcaagc cactgctggt aatgacatag 3360
tctaggtgct ctggtactgg gtgtgggttc atctgcacat gtgggtactt ctcacgatgc 3420
cggtggaaat gcactttgag gttgccacgg gtggtaaaac ggtttccaca gacattgcac 3480
ttatagggcc tctcacccgt gtgggaacga aggtggatct gcagggcact gtcactgcca 3540
aatactttgg cacagaagcg gcatttgtgc cttccaccag gcttctccaa gggacccatc 3600
acttctccgt agctcagctc accacttcca ttctttggct tcaggagccc tggggaggca 3660
gtggcctcaa ggcctcgggc tgccccaaga cactgtgctg ccagtagtcc cgtggtgctt 3720
gggaatgcca gatgaggcga ggcaatcagc tgatctgtgc tgcctggcaa ggctggggaa 3780
ggggcagggg tgggtttgtg gcttcgccca acccctccag cagagaaagg atgctgtgac 3840
cccagtgggt ggtaaaggtg gaagaaggcc tgcttgggcg tttctgcccc tgaagaggaa 3900
gaggaggagg aggaggaaga tgccagtgtc ttgctggttt ggacaggctt gatggggctg 3960
aagaggggta gtaggggctt ggtggaagag gcagtccctg tcccaggtag ctctgaggga 4020
ctggcagggg cacccaccgt ctggcctaag gagccaagca acagcacctg cctgcagatt 4080
tgctcagtca tctgcatctg atggatctgc cgctgctgca gcacccgtag ctcttccaag 4140
atcaggggga tattcaagtg gccactgcct acccctgggg gcggaggggg tggtggagga 4200
ggaggggtg caggggtcga ttctggaggt aatggggttg ctcccagctt gggactggcc 4260
aagatcaggc ccccgcctcc cccagccgct gtacctgtgg cagcgaccag gaaatgccct 4320
ggagactcct ctcctctcct ctctgggccc caggtgggat ccgtgggcac ggaggaccca 4380
gaatctgggg ggttgctatg ctctgtgtcc atgacctgag gattattgtg accctcaggc 4440
cggggttcag aggaggccga agagttgttg gggttctcct ggcccccaat tatcaccatt 4500
acaggagggt cagtagaaca tgcgttctgg tgggcgagga attcagttgg gtcagtgaat 4560
tgtgcgcagc acttggcaca gacttggggg tgatcctcct cgctagcatc acctggggag 4620
aagacaagga gagagagcgt gggtggcgca gttgggttgg gtataccgag gctctaatta 4680
acaaggaggc cagtaaccgc tagttggggg tggggagatg agctcaccat cagggccatg 4740
cagaagtcta gagctcaggc ctgatccgtg tggacaggag acaacccggc atggggcagg 4800
ggggtgggga gggaggaggg gaggggggca agagcatgct actcccctcc tcagccaccc 4860
tcccttcccc aggccacaag cgagttcacg gaataggtgt ggggacaggg gcctacgcag 4920
agaatcatgc attttctccc acccaccgaa agtcttcgcc gcccctgcgc atccccctcc 4980
gcccccaccc ctgcccagcc cgaccgaccc taccgcacct ccgagctctg ccggctcccc 5040
gcagggcacc ccgagacgag agctcctctc ggattcgtgc gccatggttg tgggggaagt 5100
ggagggccag gtggggtggg agacaatgga tattgggatt gagggaggcg atggccgctg 5160
ggtctgcggc agcctctgca cccagcggcc cagactgcgg agatggagat cggcagcggc 5220
gggggcaggg agcagcggcg gaggggggagg ggagcgagga ggcggggaga agctggagtg 5280
agaaagcggg gagaggggag atctgggagg agctgatgag gaggggagtt tatggggagg 5340
agctgctggg gagggaggcg ggagctagag gaggcgggag aagggagcgc tagcgggggc 5400
gtgggggcgg gagctcagag ctcgggagag tttccggagg cgcagtgaca ggtgctgtga 5460
agcactgcgg gggtccacct ttcccggtcc ctggccagct ccccccatct gcagatgcct 5520
ttgcccaggc ctaccctcct ccccccgccc tcccctccta agctctaggg gcacagtggg 5580
aaacgtagcc ctgctcagtg gagcaaggcg ataggcttct cttatttttc tttggataaa 5640
ggatccgctg agcttggaaa aagtggattc cagagagggt cgtctgatct cctcagaggt 5700
ctgagggcca gaagaagagg gggagatcag aacatccact cctcaccagc acacacaccc 5760
caaaatattc gaagttttgt ctcgtctttc tcacttccat tcccacccta cccccatccc 5820
tctccacaaa agaagtttct cagggtgggc ggctgcaagg tagaatttcc caggaagtca 5880
tttcaggact ctctgcggaa cactaagccc cttcactccc cgcccctcct cccctgaat 5940
aatagctgaa tgcaggttac tccgcagatc gcccagccta cacaacacct aattcataga 6000
gtccatgctt atttaataag ccatctccta tttagtaccc tcttcctcct ctattctcct 6060
cttgcaacat tcctcacacc gtcactatta aagacagtgg gtttggggag acgctagcct 6120
gcagaggcct acggaggccc acccagctct aacctggggg ggaggggagc cctcttgaaa 6180
caatgcggta ggaactacca ggcagccctc agtgtctaaa gccctttcag ccccagcctg 6240
atttgaatgc ttagaaatag ctaacacctg ctcaccatca cagaggcagc ctcctattca 6300
gacaggataa gtaagaataa aatgcctcct ggaccaggta ttctggcatt ctctttttta 6360
ccttgaaatg agtcttaaag tgcttcccac ttcctaaaat actttctctt acatgcagga 6420
agtgaccaca agtccttggt tttgtggttt ccctgggcat cagtaaacct aaattgtttt 6480
aatcccagtt ctattcttgc ctcactgata aaactgagac atggtggtca gtcacaccat 6540
gttataccac cgtttccctc ttcataaagt ggtaatattg tagctgcagt attttactca 6600

gaaaaatatt gtggggacaa aaaattgaaa aattggacaa tattaatgtg taaaccaggt 6660
atggtggtgc acacttgtag ccccagctac ttgggaggct gaggcaggag aattgcttta 6720
gtccaggagt ttgaggctgc agtgagctgt gatcacacct gtgaataacc actgccctcc 6780
agcttcggca acatagtgag gccccattac tttaaaaaaa aaaaaaagcc gggcgcggtg 6840
gctcactgta atcccagcac tttgggaggt gggcagatca cgaggtcaga agttccagac 6900
cagcatggcc aacatgttga aaccccgtct ctactaaaaa tacaaaaatt agctgggcat 6960
gatggtacac ctttaatccc agctactggg gcagctgagc caggagaatg gcttgaaccc 7020
aggaggtgga ggttgcaggg ggctgagatc gtggcattgc actccagcct gggcaacaag 7080
agtgaaactg cgtctcaaaa aaaaaaaaaa gtctaaaaaa attaatatgt acatgtgaga 7140
tttttaaagt ttggggagtc ctgaatttaa tcaatgagat aatttacatt gtcagtagca 7200
aaataatcga agtaacctta aatacacata tactaaaatt agatctgttt tccatgttgt 7260
ttgttaatct tattaatttc tgaggtaaga tattggctaa tatcagcagc atatttcaaa 7320
ggtaggaagt cttttattgc agtgggtggg ggagctgaaa caacctattt aaaatattag 7380
taacatccac tttacttctc aacataaatt ttgcctgtgt ttttaaactt aaaacagttt 7440
actgaattat gttttgaaac ttcagataat aaggctctta gcattgtgag tcataattct 7500
gaaatggacg ggttctgtgc ttccaggcct ggacttacaa atgagggagg ggggttctat 7560
ttcagtttat ggcaagtcac agttttgtgc aatgtggttt atttttacag ataaggaaac 7620
tgaagcttgg agaggttaag tgacttttcc aagttcacac agtaattcag tgaagcaagc 7680
attcagaatt ttgactcctg tccaatgctt tctcaagcac atcaactttg tatggcttcc 7740
ctaatgctag agaaagggcc ctgtgtggct tctacctgcc atttgctccc tggccttagt 7800
cagggagagg gaatcagatg gaggctttct actgagcatt tgttaattag cattgaacat 7860
ttgatatcca gttgctgttt tgtcaagtct tctgacaaga aaagaaatcc ttttcttttc 7920
atcttctcct gggaaacact gtccccttcc ttgctcttta atgaaatgtg ctttctgatg 7980
cgtaatttga tctaagctct tctttaaggt aaatttagtc cctggtgaaa ggtgactgga 8040
tcaacagcca cctgtaagag gaaccctcca tttctcagta ctttgcactc actgcacatc 8100
ctgaaaaggg gggcaggatt cttacacaaa catgaatgaa gtcacaaatg caggaataaa 8160
ctaaactggt aatggtgtcc ctagatagca gataaggtga ggtaagctat ctccggtcaa 8220
atgcaaagtc cggggtggga ctaagacctg gacaagcttg tttaaactta tagagagctg 8280
aaatgacaaa gaaaagggaa accaggtggc ttcccttcta aatctagtgt cccatcagat 8340
tgcttcttta ggcttcagag agaactgttc gggagaacaa agagaaaaat aggtgagttg 8400
tatgtagcag ggtgatacat ttgaacagcg gttttcaaat tttgctgccc attaggatta 8460
ccagaagaga gttttaaaat ttttatgttt aggtgcagtg gtctgttcct tgtagtccca 8520
gctactctgg aggctgaggc gggaggatca cttgagctca agggtttgag actccatctc 8580
aaaatctcaa aaaaaagaag aaaaaaaaaa gaaaagaaag tttaagcaca gtgggtacct 8640
catgcctata atcctagcac ttttggaggc caaggcagga ggattgcttg aggccaggaa 8700
tttgagacca gcctgcataa catagtgaga cccccatctc tgcaaaagca accaaccaac 8760
caaacttaaa aaaaatccct gtgtccaggc cacatcccag gctaattaat tcataatccc 8820
tgaggatagg atccaggcat tagtttgata aagctcctca ggtgattcca atgagaatac 8880
aaagatggtg acacaatgat gagacccaca tggaggactg ccctttccat catacctcc 8940
accctgctcc tcacagatct tacctgagct aaacttggcc acaattggga cacagacaaa 9000
atgaactctc aatgctaaat cttcccatca ggtcccctcc ctacagtgcc cacaaccaca 9060
cattaacttc cttgtatcct ttcccagtga aaaatctgct tccatgaata gaatttgata 9120
taatttacac cttactgtaa gtttaagtga ttgcatttct ttcccaggta tgggtatctt 9180
gaagcatatt tttttctttt ttaattgata tttgagccat atttcttttt ttttctttct 9240
ttttttttt ttttttttt ttgagacgga gttttgctct cattgcccag gctggagtgc 9300
aatggcatga tctcggctca ccgcaacctc cacctcccag gttcaagcga ttctcctgcc 9360
tcagccttcc caagtagctg ggattacagg catgtgccac caagcccggc taattttgta 9420
ttttttagtag agatggggtt tctccatgtg ggtcaggctg gtctcaaact cccgacctca 9480
ggtgatctgc caacctcggc ctcccaaagt gttgggatta caggcgtgag ccatcgcgcc 9540
cggccaccat atttctaatt gtaaggtgaa aggctttgtt ctacagagtt caagcatcat 9600
ccacccatta aggctggagt gaagtggcac aatcatagct cactgcagac tctacctccc 9660
aggcttaggt gatcctccca cctcagcttc ctgaatagct gggactacag gcatgcacaa 9720
tcatgcccag ctaattaaaa tatttttttc tgtagagatg aggtttcact atgttgccca 9780
ggctgtctgg aatacctggg ctcaagggat cctcctgcct tgtccccaca aagtgctgag 9840
agtacagatg taagccactg cctctggccc acttacttat tattgacact gaacaatgct 9900

aattggtagc ttccataatt atgaattgat tctgtaacta ttgctactga ctacttctta 9960
gggaaatatc tcatcttctc ctccttactc ctctttccta aatgtagaca cataataatc 10020
ctttgcaacc cagacctact aatgtaacta tggcctatgt aacacagtag actaacaggc 10080
acaatgattg gtacacctgg tgctaagtga gaaaaagata tttgtttcca gaacaggaat 10140
atcttagatc aaacataaga atgttctttt aatgaaaatt tctttgactt caaaggactc 10200
aacacttaac atggaattca taccattttg gagctgggac ttcagagatc tgacactctc 10260
attgtcattg tgcacagtga ttcagacctg agttaaagtc ccagctctag aacattctaa 10320
tatttgtgat cttgggaaaa tttcttaatc tctcccagag tttgttttct tatttttttt 10380
tgggacagag tttcactctt gttgcccagg ctggagtgca atggcacgat cttagctcac 10440
cgcaacctcc gcctcccagg ttcaagcgat tctcctgcct caccctccct agtagctggg 10500
attacaggca tgtgccacca cgcccggcta attttgtatt tttttagtag agacggggtt 10560
tcttcatatt gctcaggctg gtctcaaact cccagcctca ggtgatctgc ccacctcggc 10620
ctcccaaagt gctggattac aggcatgagc caccgcgcct ggccagcctt ttttttttt 10680
ttgagacgga gtctcgctct gtcgcccagg ctggagtgca atggtcgccc aggctggagt 10740
gcaatggtgt gatctcggct cactgcaatc tccgcctcct gggttcaaac gattttcctg 10800
cctcagcctc ccaagtagct gggattacag gtgtgcgcca tcacacccag ctaatttttg 10860
tatttttagt agagatgagg tttcaccttg ttggccaggc tggtcttgaa ctcctgacct 10920
caagtgattt gcccacctca gcctcccaaa gtgctgagat tacaggcatg agctgctgtg 10980
cccggctgat ttctcttctt taaaatgagg gtactgccat acaaaggaag gaaattctga 11040
tacatgctac aacatgaatg aactttgtaa acattatgct ttcagacaaa tttgacttta 11100
attgagaaaa aaagagaaaa catactaagt gcaataaagc agacacaaaa ggacaaatat 11160
tgtatgattc cattagtatg aggtacccaa acattatatg agtccattaa tatgaaattt 11220
ggcaaggtca cacatacaga aagcagagta gaggctaaca gggctaaggg aatgggagaa 11280
tggggattta ttgtttaacg gttacagttt ctgtttgatg atgaaaaaga tattgaaaca 11340
gcagtaatgg ttacataaca tagtgaatgt acttaatgcc actgaattgt acacttaaaa 11400
atggttaaaa tggtaaattt tattacacat attttacaat aaaaaaattt tagccaggtg 11460
tggtggcatg cacctgtaat cccagctgtt caggaggctg aggcaggaga atctcttgaa 11520
ccctggaggt ggaggtttca gtgagccgag acgtgccact gcactccagc ctgggcaaca 11580
gagtaggact tggtctcaaa aaaagaaaaa aatttttttt gtaataataa gggagttggg 11640
gctgggcgtg gtggctcacg cctgtaatcc cagcactttg ggaggccaaa gtgggcggat 11700
catgaagtca ggagatcgag accatcctgg ctaacacagt gaaaccctgt ctctactgaa 11760
aatataaaaa attagccagg tgtggtggcg ggcgcctgta gtcccagcta cttgggaggc 11820
tgaggcagga gaatggtgtg aacccgggag gcggagctcg cagtgagcca agatcgcgcc 11880
actgcacccc agcctgggcg acagagcgag actccgtctc aaaataataa aaataaataa 11940
ataaataaat aaaataataa taataacgga gttgggagga aaaagaggaa atgcaaaaag 12000
ggcctagcac agtacctgaa tgctccacaa atattagcca tgggtgttag ttattatttg 12060
aatgtcaaaa gctgaatgaa gccctggggt aagaaaggtc acatgtgccc aaggtcacat 12120
agcttcaagg tccacactag attgaaaacc aagttttctg ttttcttatc tagtactctg 12180
taacaccagg actgagatac tctctattcc aaaatgtgtt ttttctgatc tgggaatacc 12240
taggttgagt ggcccaggga tcaataacct gagagatgag gctctttact tccaaatgta 12300
aacagagccc ccaaaactct acctttgcct tctttcctct cttgctgttc ttgctatctg 12360
ccaacttcca tctaaagtac tccctctct ccctctagat ctgtttggct gctgtcctgg 12420
tttcttcttc tcactaaata tctgggtttc tgattgtttc ctttatttcc cagatgtact 12480
ggtttgcatt tttcccccag tcacatcctt tgtgttctct aatccagatt tctagactct 12540
gtaggggaga gagaaggttt tttttttcc tctctagagt ttttaagtga atagagtatt 12600
tcctgcccat cacttatatg caataactgt tctgttaggt tttgatgctc tggttaggga 12660
agctgagcaa aaacggctgg aaaacagatt tttcagactg tttcttggtg atgtcttagg 12720
tcactgcaga attttggctt ttaaaatatg taacaaaggc tcagcatttg catgttgtat 12780
atggcacata ttgcttacaa gaaggcaaaa gactcctgga aacattactg gcaccctaga 12840
ctactgacta aatgtcttct gatactcatg atgatatcca taatttcaca ggtacaccaa 12900
aggatacatg tgcccctaaa taagagccct tcctccctaa ctgtggagca tgctctgggg 12960
tagaaggaag tcagatgcct gaagatcaca taagtgaata gaaaccctgt ctataaaaaa 13020
ttagggaaaa ggagagctct cattctgttt tgcagaatgg atgctgcccc attcatgatt 13080
aagaaaattt attaatttaa aagaaaacca gaaaatgtga aatttatata ttataagctt 13140
ataagatcca ggaggaattt tagatacgat caaatagagc cacctcattt tgcagatgag 13200

gcccaatgac atccagatca taagtagcct aggatctttc actccagggg aattctgatg 13260
agaaaatcct taggctttct tacggtagat cttaacagag ggtgctactg cttccttgct 13320
ccttacattt gttcctgcct ttcatagctc aaaggcaaat tttcatcaaa aatttgttga 13380
tgccattggg tttaaacctt tactgtttct atggggatgg ctttgtaaca gcattaccat 13440
gcccccaggt ggaagctata tcttaaaggg cttgaaaatc cattcaagac agccgctaaa 13500
gatagctttt gactccctca cagaagattt ttcctcagct atgatatggg gaatgggtga 13560
gcagatggga gaagtaggaa gaagaggaga gaatgcttct tgggggtttg gaggggtgtt 13620
cagcatagtt ccacaatcaa accagcagga gagcagaact gtgaggcaac tctggggagg 13680
agttgaggct ctaggggaag tctcctgtag agcacaagca ggaaacatcc ggcctatagc 13740
agcattaaga agggctaatg tgtctcagga gggaaggatg ccatcaccat agaacctcta 13800
aatatgggca cagtaggatc ccagaaaagc agtgtttcgg ggaggatgcg ttctgcccaa 13860
aacatgtctg ttaaggttat tttgtagcac atggagcgct gatttgacct caagtttttg 13920
ttttttaaca ggtggaaagg caagtttaat ctacaatttt agtcgccacc aatacactct 13980
cttagagctt ttcatgacac gtctcataaa gaaatgctga tggccgggag cggtggctca 14040
cgcctgtaat cccagcactt tgggaggcca aggcgggcag attacgagat caggagatcc 14100
agagcatcct ggctaacacg gtgaaacccc gtctctacta aaaatacaaa aaattagccc 14160
ggcgtggtgg caggcgccta tagttccagc tactcgggag gctgaggcag gagaatggcg 14220
tgaacctggg aggtggagcg ggcagtgagc caagattgca ccactgcact ccaacctggg 14280
cgacagagcg agactctctc tcaaaaaaaa aaaaagaaaa aaagaaaaag aaaagaaaag 14340
aaaaaaaaag aaatgctgac gtttgccaag aggttcctga gttttggtca tactacagca 14400
cttgcaggca gtgtcactgc attcacatat aatgataata acgatattca cacatattaa 14460
gcacttattt atgctaggta tttttccaag ggatttacac atattaactc atttagattt 14520
tcacaacaac ctaatgaggt agctagtata cacatcttta tttcacagat gaggaaactg 14580
aagcatagag aggcaaaata aaccagccaa ggtcacatag ctaaccaagt ggtggagctg 14640
ggatttgtct aaaagtctgg tttcagaacc cttgtgctta atcctatact atactgttgg 14700
gtgtatcaac tgtatgctaa acagttgcct gtctggagcc aggacttcca gactttcagt 14760
ctgcacatat ggagccatac cactgacaag tatgtccaaa acttctttga tcctaagaat 14820
tacctggaca attgcaaaat atatagattc ccacaccctg gctcagatgt actcacaatc 14880
aggcaagttt ttaaaaccca ggtttagtgg gtttagtgag cactaccagc cagccctgag 14940
cattaggaaa ttgaagtttt tgtcctgatt ttgcttctgt ctctcagact ctgagcaatt 15000
tcactcttca attccctgct tgctctactg tctgcctgtc acttaacgga atgttacaag 15060
aatacataca atttttcccc ctcataaggg acacctgttg cttcaaaaac acggtatcct 15120
cataaaatga tatgcatgta gtaacaggtg tattttcttg cacttctttt gttttgtttt 15180
gttttttgct ttgctttcct tgaagcacaa acctaagccc ctcatccaga cctagccttc 15240
agctgtcctc caggtgacac gcatacacac cccaaaccag gctgcattct gaccgacctt 15300
agctctctcc ctctgggagc tctgatcggc tctcagttca gcccaacaat gagaaacttt 15360
tttctcgtct ccctcagggg agccttcacg tttatccaat tcattctctt gcaacccaac 15420
tctccagaaa gaaaaggggg gaaaatccca ccccgaagag acggtcttca ggtctgagga 15480
cgttacttag caacggcaca aagaccagtg agcaaaggga gacctgagga gaaaactctt 15540
gggtggggag acagagccag tttgaaaact ccatttcatc cagagaaaaa caaggaaaac 15600
acaaacagaa tcaatcccaa gtaacaagcg gggcttctcc ccagcgcagg tcatctctta 15660
ctccctgcat ctcaactcct tcaaaccccc agtgaccaag tccgcccccg cctggtttcg 15720
cccatggccc gagtgccctc cccttgccct ggcctgaccc acacaggctt ggacttaggg 15780
gcccccaccc ctccccaggc acccaccgtt ctcagacgcg ctgggacctt cgcagtccga 15840
gattaactgt tggggtttcc gctgctttcg ccgagacatt cccgggtaga gagttgggag 15900
agggaggggc aacgctcact tggtcttaac cggggtgacc tggtctcgtc tccccccttgg 15960
gtccgaagcc aattgatgcc tctcccccag cgcaaatcac tgtgaagcag agatgttctt 16020
ctttcccaga gacacagact ctctctctct ctctgattct ctgttcttga ctctctctct 16080

<210> 3
<211> 1002
<212> PRT

<213> Mus musculus

<400> 3

```
Met Ala Gln Glu Thr Gly Ser Ser Ser Arg Leu Gly Gly Pro Cys Gly
 1               5                  10                  15
Glu Pro Ala Glu Arg Gly Gly Asp Ala Ser Glu Glu His His Pro Gln
            20                  25                  30
Val Cys Ala Lys Cys Cys Ala Gln Phe Ser Asp Pro Thr Glu Phe Leu
        35                  40                  45
Ala His Gln Asn Ser Cys Cys Thr Asp Pro Pro Val Met Val Ile Ile
    50                  55                  60
Gly Gly Gln Glu Asn Pro Ser Asn Ser Ser Ala Ser Ser Ala Pro Arg
65                  70                  75                  80
Pro Glu Gly His Ser Arg Ser Gln Val Met Asp Thr Glu His Ser Asn
                85                  90                  95
Pro Pro Asp Ser Gly Ser Ser Gly Pro Pro Asp Pro Thr Trp Gly Pro
            100                 105                 110
Glu Arg Arg Gly Glu Glu Ser Ser Gly Gln Phe Leu Val Ala Ala Thr
        115                 120                 125
Gly Thr Ala Ala Gly Gly Gly Gly Gly Leu Ile Leu Ala Ser Pro Lys
    130                 135                 140
Leu Gly Ala Thr Pro Leu Pro Pro Glu Ser Thr Pro Ala Pro Pro Pro
145                 150                 155                 160
Pro Pro Pro Pro Pro Pro Pro Pro Gly Val Gly Ser Gly His Leu Asn
                165                 170                 175
Ile Pro Leu Ile Leu Glu Glu Leu Arg Val Leu Gln Gln Arg Gln Ile
            180                 185                 190
His Gln Met Gln Met Thr Glu Gln Ile Cys Arg Gln Val Leu Leu Leu
        195                 200                 205
Gly Ser Leu Gly Gln Thr Val Gly Ala Pro Ala Ser Pro Ser Glu Leu
    210                 215                 220
Pro Gly Thr Gly Ala Ala Ser Ser Thr Lys Pro Leu Leu Pro Leu Phe
225                 230                 235                 240
Ser Pro Ile Lys Pro Ala Gln Thr Gly Lys Thr Thr Ala Ser Ser Ser
                245                 250                 255
Ser Ser Ser Ser Ser Ser Gly Ala Glu Pro Pro Lys Gln Ala Phe Phe
            260                 265                 270
His Leu Tyr His Pro Leu Gly Ser Gln His Pro Phe Ser Val Gly Gly
        275                 280                 285
Val Gly Arg Ser His Lys Pro Thr Pro Ala Pro Ser Pro Ala Leu Pro
    290                 295                 300
Gly Ser Thr Asp Gln Leu Ile Ala Ser Pro His Leu Ala Phe Pro Gly
305                 310                 315                 320
Thr Thr Gly Leu Leu Ala Ala Gln Cys Leu Gly Ala Ala Arg Gly Leu
                325                 330                 335
Glu Ala Ala Ala Ser Pro Gly Leu Leu Lys Pro Lys Asn Gly Ser Gly
            340                 345                 350
Glu Leu Gly Tyr Gly Glu Val Ile Ser Ser Leu Glu Lys Pro Gly Gly
        355                 360                 365
Arg His Lys Cys Arg Phe Cys Ala Lys Val Phe Gly Ser Asp Ser Ala
    370                 375                 380
Leu Gln Ile His Leu Arg Ser His Thr Gly Glu Arg Pro Tyr Lys Cys
385                 390                 395                 400
Asn Val Cys Gly Asn Arg Phe Thr Thr Arg Gly Asn Leu Lys Val His
                405                 410                 415
```

```
Phe His Arg His Arg Glu Lys Tyr Pro His Val Gln Met Asn Pro His
                420                 425                 430
Pro Val Pro Glu His Leu Asp Tyr Val Ile Thr Ser Ser Gly Leu Pro
        435                 440                 445
Tyr Gly Met Ser Val Pro Pro Glu Lys Ala Glu Glu Glu Ala Gly Thr
    450                 455                 460
Pro Gly Gly Gly Val Glu Arg Lys Pro Leu Val Ala Ser Thr Thr Ala
465                 470                 475                 480
Leu Ser Ala Thr Glu Ser Leu Thr Leu Leu Ser Thr Gly Thr Ser Thr
                485                 490                 495
Ala Val Ala Pro Gly Leu Pro Thr Phe Asn Lys Phe Val Leu Met Lys
            500                 505                 510
Ala Val Glu Pro Lys Ser Lys Ala Asp Glu Asn Thr Pro Pro Gly Ser
        515                 520                 525
Glu Gly Ser Ala Ile Ala Gly Val Ala Asp Ser Gly Ser Ala Thr Arg
    530                 535                 540
Met Gln Leu Ser Lys Leu Val Thr Ser Leu Pro Ser Trp Ala Leu Leu
545                 550                 555                 560
Thr Asn His Leu Lys Ser Thr Gly Ser Phe Pro Phe Pro Tyr Val Leu
                565                 570                 575
Glu Pro Leu Gly Ala Ser Pro Ser Glu Thr Ser Lys Leu Gln Gln Leu
            580                 585                 590
Val Glu Lys Ile Asp Arg Gln Gly Ala Val Ala Val Ala Ser Thr Ala
        595                 600                 605
Ser Gly Ala Pro Thr Thr Ser Ala Pro Ala Pro Ser Ser Ser Ala Ser
    610                 615                 620
Gly Pro Asn Gln Cys Val Ile Cys Leu Arg Val Leu Ser Cys Pro Arg
625                 630                 635                 640
Ala Leu Arg Leu His Tyr Gly Gln His Gly Gly Glu Arg Pro Phe Lys
                645                 650                 655
Cys Lys Val Cys Gly Arg Ala Phe Ser Thr Arg Gly Asn Leu Arg Ala
            660                 665                 670
His Phe Val Gly His Lys Thr Ser Pro Ala Ala Arg Ala Gln Asn Ser
        675                 680                 685
Cys Pro Ile Cys Gln Lys Lys Phe Thr Asn Ala Val Thr Leu Gln Gln
    690                 695                 700
His Val Arg Met His Leu Gly Gly Gln Ile Pro Asn Gly Gly Ser Ala
705                 710                 715                 720
Leu Ser Glu Gly Gly Gly Ala Ala Gln Glu Asn Ser Ser Glu Gln Ser
                725                 730                 735
Thr Ala Ser Gly Pro Gly Ser Phe Pro Gln Pro Gln Ser Gln Gln Pro
            740                 745                 750
Ser Pro Glu Glu Glu Met Ser Glu Glu Glu Glu Glu Asp Glu Glu Glu
        755                 760                 765
Glu Glu Asp Val Thr Asp Glu Asp Ser Leu Ala Gly Arg Gly Ser Glu
    770                 775                 780
Ser Gly Gly Glu Lys Ala Ile Ser Val Arg Gly Asp Ser Glu Glu Val
785                 790                 795                 800
Ser Gly Ala Glu Glu Glu Val Ala Thr Ser Val Ala Ala Pro Thr Thr
                805                 810                 815
Val Lys Glu Met Asp Ser Asn Glu Lys Ala Pro Gln His Thr Leu Pro
            820                 825                 830
Pro Pro Pro Pro Pro Pro Asp Asn Leu Asp His Pro Gln Pro Met Glu
        835                 840                 845
Gln Gly Thr Ser Asp Val Ser Gly Ala Met Glu Glu Glu Ala Lys Leu
```

```
    850                 855                 860
Glu Gly Ile Ser Ser Pro Met Ala Ala Leu Thr Gln Glu Gly Glu Gly
865                 870                 875                 880
Thr Ser Thr Pro Leu Val Glu Glu Leu Asn Leu Pro Glu Ala Met Lys
                885                 890                 895
Lys Asp Pro Gly Glu Ser Ser Gly Arg Lys Ala Cys Glu Val Cys Gly
            900                 905                 910
Gln Ser Phe Pro Thr Gln Thr Ala Leu Glu Glu His Gln Lys Thr His
        915                 920                 925
Pro Lys Asp Gly Pro Leu Phe Thr Cys Val Phe Cys Arg Gln Gly Phe
    930                 935                 940
Leu Asp Arg Ala Thr Leu Lys Lys His Met Leu Leu Ala His His Gln
945                 950                 955                 960
Val Pro Pro Phe Ala Pro His Gly Pro Gln Asn Ile Ala Thr Leu Ser
                965                 970                 975
Leu Val Pro Gly Cys Ser Ser Ser Ile Pro Ser Pro Gly Leu Ser Pro
            980                 985                 990
Phe Pro Arg Lys Asp Asp Pro Thr Met Pro
        995                 1000
```

<210> 4
<211> 4547
<212> DNA
<213> Mus musculus

<400> 4

```
atggcgcagg aaaccgggag cagctctcga ctcgggggac cctgcgggga gcctgcggag 60
cgcggaggtg atgctagcga ggaacaccac ccccaagtct gtgccaaatg ctgcgcacaa 120
ttttctgacc cgaccgaatt cctcgctcac cagaactcat gttgcactga cccaccggta 180
atggtgataa ttggaggcca ggagaatccc agcaactctt cagcctcctc tgcgccccga 240
ccagagggcc acagtaggtc ccaggtcatg gatacagagc acagcaatcc cccagattct 300
gggtcctctg ggcccccgga tcccacttgg gggccagagc ggaggggaga ggaatcttct 360
gggcaattcc tggtcgctgc cacaggtaca gcggctgggg gaggtggggg ccttatcttg 420
gccagtccca agctgggagc aaccccatta cctccagaat ccactcctgc acccccctcct 480
cccccaccac ccccтccccc tccaggtgta ggcagtggcc acttgaacat tcctctgatc 540
ttggaagagt tgcgggtgct gcagcagcgc cagattcacc agatgcagat gactgaacaa 600
atctgccgcc aggtgctgct acttggctcc ttggggcaga ccgtgggtgc ccctgccagt 660
ccctcagagc tacctgggac aggggctgcc tcttccacca agcccctcct gcctctcttc 720
agtcccatca agccagcgca aactggcaag acactggcat cttcctcttc gtcatcctcc 780
tcctctggag ctgaaccgcc taagcaggct ttcttccacc tttaccatcc actgggatca 840
cagcatcctt tctctgtagg aggggttggg cggagccaca aacccacccc tgccccttcc 900
cctgcgctgc caggcagtac ggatcagctg attgcttcac ctcatctggc attcccaggc 960
accactggac tcctggcagc tcagtgtctt ggggcagcaa ggggccttga ggctgctgcc 1020
tccccagggc tcctgaagcc aaagaacgga agtggtgaac tgggctatgg ggaagtgatc 1080
agttccttgg agaaacccgg tggaaggcac aaatgccgct tttgtgcaaa agtattcggc 1140
agtgacagcg ccctgcagat ccaccttcgt tcccacactg gtgagaggcc ctataagtgc 1200
aacgtctgtg gtaaccgttt cacaactcgg ggcaacctca aagtacattt tcaccggcat 1260
cgtgagaagt acccacatgt gcaaatgaat ccacatccag taccggagca cctagactac 1320
gtcatcacca gcagtgggct gccttacgga atgtctgtgc caccagagaa agcagaagag 1380
gaggcaggca caccaggcgg aggtgttgaa cgcaaacccc tagtggcctc caccacagca 1440
ctcagtgcca cagagagcct gacactgctc tccactggca caagcacagc agtggctcct 1500
gggctcccta ctttcaacaa gtttgtgctc atgaaggcag tggaacccaa gagtaaagcg 1560
gatgagaaca cgcccccagg gagtgagggc tccgccatcg ctggagtagc agacagtggc 1620
tcagcaaccc gaatgcagct aagtaagctg gtgacgtcac taccgagttg ggcactgctt 1680
```

actaatcact tgaagtcaac tggaagtttc cccttccctt atgtgctaga acccttgggg 1740
gcttcgcctt ctgagacctc aaagctgcag cagctagtag aaaagattga ccgccaagga 1800
gctgtggcgg tggcatctac tgcctcggga gctcccacca cttctgcccc tgcaccttcc 1860
tcctccgctt ctggacctaa ccagtgtgtg atctgtcttc gggtcctgag ctgccctcgg 1920
gctctacgcc tgcattatgg ccaacatgga ggtgagcggc ccttcaagtg taaagtgtgt 1980
ggccgagctt tctccacaag gggcaatttg cgcgcacatt tcgtgggtca caagaccagt 2040
ccagctgccc gggctcagaa ctcctgcccc atttgtcaga agaagttcac taatgctgtc 2100
actctgcagc aacatgttcg gatgcacctg gggggccaga tccccaatgg gggttccgca 2160
ctttctgaag gtgggggagc tgcccaggaa aacagctctg agcagtctac agcctctgga 2220
ccagggagtt tcccccagcc gcagtcccag cagccatctc cagaagagga gatgtctgag 2280
gaagaggaag aggatgagga agaggaggaa gacgtgacag atgaagattc cctagcagga 2340
agaggctctg agagtggggg agagaaggcc atatcagtac gaggtgactc agaagaggta 2400
tctggggcag aggaagaagt ggcaacatca gtagcagcac ccaccactgt gaaggagatg 2460
gacagtaatg agaaagcccc tcaacacact ctgccgccac ctccgccacc acccgacaac 2520
ctggatcatc cccaacccat ggagcaggga accagtgatg tttccggagc catggaggaa 2580
gaagccaaac tggagggaat ctcaagcccg atggcagccc tcacccaaga aggggagggc 2640
accagcaccc ctttggtgga agagctgaac ttaccggaag ccatgaagaa ggatccagga 2700
gagagcagcg gcaggaaggc ctgtgaagta tgtggccaga gctttcctac ccagacagct 2760
ctggaggagc atcagaagac ccatcccaag gatgggccac tcttcacttg tgtcttctgc 2820
aggcagggct tccttgaccg tgctaccctc aagaagcaca tgctgttggc tcaccaccag 2880
gtaccgccct ttgcacccca tggccctcag aatattgcta ctctttcctt ggtccctggc 2940
tgttcctcct ccatcccttc tccagggctc tccccattcc ctcgaaaaga tgaccccacc 3000
atgccatgag cctgctttct gtacctggtc ctctatgacc cagagagcag aaacctgaga 3060
gcttcataga ggaactccaa gatttactca ccctcctctt gtcctttctc aagtcctgac 3120
atgatgtttc tagtggcttc ttctctagtc cctgagcttg acaattgcct ttgaaagaga 3180
atgtcccctt aagaaatttt tatcaccttt ttgttctgtg taactaaggg aaacaaattc 3240
cctatagctt ttacattctc aagggggagc tctctcctct tctccctttc cctttggcag 3300
gtatactaga acccccatcc ttggagtggc agccttggtc caaggggctg gcaactgtcc 3360
atggaaggcc cagcgttact ccttggtgat cttgaccacc ctgcaagact ttctagggcc 3420
gggaccttct tgagaagctt gtaaggggtg gtaggtttct ttctgcaacc actacccagt 3480
tttccactga gccctggagt tctggaccta cctgcattgc cactcgggcc ctagtaccat 3540
cattgctgtg aaagcccagg aactgtgttt cacaaggtga ctccagtgac atgatccaga 3600
gaggcaaaga acatagcctc cggaagttga ggctgtgccc aacaagcaca ccggaagaaa 3660
gaagaaacta taacttcttt ctccttcccc cctgctccag agagtgctgg caataaagat 3720
attctagcaa ttggtgactc accctagaag gtagggacaa gtgaaggact gggacccttt 3780
ttgcagtatg ttccttgact cgccacattg aggcaaagat agtggctggt caagatgcca 3840
ggactactcc agcttcccat catgtcctct caaccaacaa gcaggtttcc taccaagagg 3900
tctctcgtgt gatagtttag ggagtatgaa gtttctaact ctaaagaatc ctgttggtga 3960
ggatgattat ttaagcaatg atggggagtt gagggttgtt gctaaaacag gcattgctgg 4020
gaatctattt gatgaagaac aggacttgat gtaaggggac tcgatgttca gctcttgtga 4080
gtatgaacgt tttctttgag ctaatggtga tgtggtatgc agaggtacca ggggccatgg 4140
gggtgtgtgt gcttcctgtc actagaatgt ttttagtttt agatgactcc ctattttatt 4200
ccctcacccc ttgtatttcc cttgctgtct tctcaaaacc cctttcctcc cccagttttg 4260
cctgaccatg ggccagagct tatgtcttat ttttttttcta gaagttgaga gacagagctt 4320
caagtggttt ccccccgtct ctgtcttgta gtgagatgta gtatttactc ttaacatagg 4380
atcctgtgga acaggtgttc tgagaagact gaattttgct gttagctgtt gtcaatgatg 4440
attctctaaa gtagtgggct ccagagctcc ctaacacagt gaaatgtgta agagccgaga 4500
ggggagatac tagaattttt tccttcatca ttaaaggtgt tttggct 4547

<210> 5
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> derived from human Sal2 gene

<400> 5
ccacaaccat ggcgaatccg ag 22

<210> 6
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from human Sal2 gene

<400> 6
ggtgatggaa ggcgaacagc cagg 24

<210> 7
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from Human Sal2 gene

<400> 7
cttgttaatt agagcctcgg tatacc 26

<210> 8
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> derived from Human Sal2 gene

<400> 8
gcacggagga cccagaatct gg 22

<210> 9
<211> 63
<212> DNA
<213> Polyoma virus

<400> 9
gatatacttt gtaatgtgca agaaggcgac gacccttga aggacatatg tgaatatagc 60
tga 63

<210> 10
<211> 20
<212> PRT
<213> Polyoma virus

<400> 10
Asp Ile Leu Cys Asn Val Gln Glu Gly Asp Asp Pro Leu Lys Asp Ile
1 5 10 15

Cys Glu Tyr Ser
20

<210> 11
<211> 19
<212> PRT
<213> TMD25 mutant Polyoma virus

<400> 11
Asp Ile Leu Cys Asn Val Gln Glu Asp Phe Val Met Cys Lys Lys Ala
1 5 10 15
Thr Thr Pro

<210> 12
<211> 60
<212> DNA
<213> TMD25 mutant Polyoma virus

<400> 12
gatatacttt gtaatgtgca agaagacttt gtaatgtgca agaaggcgac gacccttga 60

<210> 13
<211> 16
<212> PRT
<213> Polyoma virus

<400> 13
Asn Val Gln Glu Gly Asp Asp Pro Leu Lys Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 14
<211> 14
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 14
Asn Val Gln Glu Gly Asp Asp Pro Leu Lys Asp Ile Cys Glu
1 5 10

<210> 15
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 15
Asn Val Gln Glu Gly Asp Asp Pro Leu Lys
1 5 10

<210> 16
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 16
Asn Val Gln Glu Gly Asp Asp
1 5

<210> 17
<211> 4
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 17
Asn Val Gln Glu
1

<210> 18
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 18
Asn Val Gln Glu Gly Asp Asp Leu Lys Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 19
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 19
Asn Val Gln Glu Gly Asp Asp Pro Lys Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 20
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 20
Asn Val Gln Glu Gly Asp Asp Pro Leu Asp Ile Cys Glu Tyr Ser
1 5 10 15

<210> 21
<211> 13
<212> PRT
<213> Artificial Sequence

<220>
<223> dervived from Polyoma virus large T antigen

<400> 21
Asn Val Gln Glu Gly Asp Asp Asp Ile Cys Glu Tyr Ser
1 5 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/21354

A. CLASSIFICATION OF SUBJECT MATTER
IPC(7) : C12Q 1/68; C12N 5/00; C07H 21/02, 21/04
US CL : 435/6, 325; 536/23.1, 24.5; 935/1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6, 325; 536/23.1, 24.5; 935/1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EAST, MEDLINE, BIOSIS, CAPLUS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FREUND et al. Host range and cell cycle activation properties of polyomavirus large T-antigen mutants defective in pRB binding. J Virology. November 1994, vol. 68, no. 11, pages 7227-7234, see entire document.	1, 2, 7-11
A	US 6,214,544 B1 (FISHER) 10 APRIL 2001, SEE ABSTRACT AND CLAIMS.	1-11

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified).

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
15 November 2001 (15.11.2001)

Date of mailing of the international search report
13 FEB 2002

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703)305-3230

Authorized officer
Q. JANICE LI for Janice Li
Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US01/21354

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-11

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/21354

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional examination fees must be paid.

Group I, claims 1-11, drawn to a method identifying a cellular protein involved in the susceptibility to proliferative disease.

Group II, claim 12, drawn to a tumor host range virus.

Group III, claims 13-19, drawn to a method of determining gene alteration comprising determining whether a cell can act as a permissive host for the propagation of a T-HR mutant.

Group IV, claims 20-30, drawn to a method for cancer cell killing.

Group V, claims 31-42, drawn to a method for risk assessment of a proliferative disease comprising detecting alteration of a *Sal2* nucleic acid in a mammal.

Group VI, claims 43-48, drawn to a method for risk assessment of a proliferative disease comprising detecting alteration of a *Sal2* protein in a mammal.

Group VII, claims 49-52, 54, 59, 60, drawn to a knockout mouse comprising a mutation in a genomic mSal2 gene.

Group VIII, claims 53, 55-58, 61, drawn to a transgenic mouse whose genome comprising a nucleic acid construct including a Sal2 gene.

Group IX, claims 62-69, 72, 74, 75, drawn to a method of identifying a compound which alters cell proliferation by measuring cell proliferation of two different abnormal test cells.

Group X, claims 70, 71, 73, 76, drawn to a method of identifying a compound which alters cell proliferation by measuring cellular *Sal2* levels.

The inventions listed as Groups I-X do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions listed as Groups I and II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I is drawn to a method of identifying a cellular protein, whereas group II is drawn to a tumor host range virus. Thus, they lack the same special technical feature.

The inventions listed as Groups III-VI, IX, X, and I do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I, III-VI, IX, X are drawn to different methods, such as for identifying a cellular protein, assessing a gene alteration or cell killing. The different methods using different test criteria, different starting materials, have different method steps, different mode of operation, and different technical considerations. Thus, they lack the same special technical feature.

The inventions listed as Groups VI and V do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups VI differs from group V in that they are different methods for risk assessment, the different method using different test criteria, i.e. a Sal2 nucleic acid or a Sal2 protein, the method steps and mode of operation for identifying a nucleic acid differs from that of a protein. Thus, they lack the same special technical feature.

The inventions listed as Groups X and IX do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups X differs from group IX in that they are different methods for identifying a compound, the different method using different test criteria, i.e. state of cell proliferation or levels of Sal 2 expression, the method steps and mode of operation for the two differs significantly that they lack the same special technical feature.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/21354

The inventions listed as Groups VIII and VII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups VIII differs from group VII in that they are different products, a Sal2 transgene mouse and a Sal2 gene knockout mouse differs in their genomic sequence structures and phenotype . Thus, they lack the same special technical feature.